Journal of Multivariate Analysis 154 (2017) 112–134

Contents lists available at ScienceDirect

Journal of Multivariate Analysis

journal homepage: www.elsevier.com/locate/jmva

Nonparametric estimation of a latent variable model

Augustin Kelava [a], Michael Kohler [b], Adam Krzyżak [c,*], Tim Fabian Schaffland [a]

[a] *Wirtschafts-und Sozialwissenschaftliche Fakultät, Hector-Institut für Empirische Bildungsforschung, Universität Tübingen, Europastraße 6, 72072 Tübingen, Germany*

[b] *Fachbereich Mathematik, Technische Universität Darmstadt, Schloßgartenstraße 7, 64289 Darmstadt, Germany*

[c] *Department of Computer Science and Software Engineering, Concordia University, 1455, boul. de Maisonneuve ouest, Montréal, Québec, Canada H3G 1M8*

ARTICLE INFO

Article history:
Received 17 April 2015
Available online 1 November 2016

AMS subject classifications:
primary 62G08
secondary 62G20

Keywords:
Common factor analysis
Latent variables
Nonparametric regression
Consistency

ABSTRACT

In this paper a nonparametric latent variable model is estimated without specifying the underlying distributions. The main idea is to estimate in a first step a common factor analysis model under the assumption that each manifest variable is influenced by at most one of the latent variables. In a second step nonparametric regression is used to analyze the relation between the latent variables. Theoretical results concerning consistency of the estimates are presented, and the finite sample size performance of the estimates is illustrated by applying them to simulated data.

1. Introduction

Latent variable models provide statistical tool for explaining and analyzing underlying structure of multivariate data by using the idea that observable phenomena are influenced by underlying factors which cannot be observed or measured directly. They have applications in various areas including psychology, social sciences, education or economics, where theoretical concepts such as intelligence, desirability or welfare cannot be measured directly but instead observable indicators (or manifest variables) are given.

One possibility to fit latent variable models to data is to assume that the underlying distribution is Gaussian, and therefore it is uniquely determined by its covariance structure. Then the maximum likelihood principle together with structural assumptions on the underlying latent variable model can be used to fit the latent variable model to observed data.

In contrast in this paper we try to avoid any assumption on the class of the underlying distributions. Given multivariate random variables X and Y, we approximate them by linear combinations of suitable latent variables Z_1 and Z_2 and then use nonparametric regression to study the relation between Z_1 and Z_2. In this way the whole procedure splits into two separate problems: In a first step we fit a common factor analysis model to X and Y. And then we apply suitable nonparametric regression techniques to analyze the relation between the latent variables in this model.

* Corresponding author.
E-mail addresses: augustin.kelava@uni-tuebingen.de (A. Kelava), kohler@mathematik.tu-darmstadt.de (M. Kohler), krzyzak@cs.concordia.ca (A. Krzyżak), tim.schaffland@gmail.com (T.F. Schaffland).

http://dx.doi.org/10.1016/j.jmva.2016.10.006

The main trick in estimation of the common factor analysis model is to estimate the values of (Z_1, Z_2) in such a way that the corresponding empirical distribution asymptotically satisfies the conditions that characterize the distribution of (Z_1, Z_2) uniquely. This primarily requires independence of (Z_1, Z_2) of the random errors occurring in the manifest variables, and we ensure this by minimizing some kind of distance between the empirical cumulative distribution function of all these random variables and the product of the marginal cumulative distribution functions.

Our main theoretical result is that the empirical distribution of the estimated values of (Z_1, Z_2) converges weakly, with probability 1, to the distribution of (Z_1, Z_2). We use this result to define the least squares estimates of the regression function of (Z_1, Z_2). We show that our regression estimate is strongly consistent whenever the regression function is Lipschitz-continuous and bounded. The finite sample size performance of our estimates is illustrated by applying them to simulated data.

1.1. Discussion of related results

Surveys on latent variables and its applications can be found, e.g., in [1,2].

One way to determine latent variable models is the use of principal component analysis; see, e.g., Section 14.5 in Hastie, Tibshirani and Friedman [3]. There the manifest variables are approximated by the best linear approximation of a given rank. The obvious drawback is that in this case the sum of the latent variable and its random error is approximated. The classical factor analysis model takes into account these random errors. If we assume that all random variables are Gaussian, then the model can be fitted by maximum likelihood; see, e.g., Section 14.7 in Hastie, Tibshirani and Friedman [3]. In the independent component analysis (described, e.g., in [4]) the latent variables are assumed to be independent, which resolves any identifiability problem in the above approaches. However, this assumption is often not realistic in applications and cannot be used in the context of regression estimation. Identifiability conditions for latent parameters in hidden Markov models and random graph mixture models have been discussed in [5–7].

Independent factor analysis model which is often used for dimensionality reduction assumes that random variables are generated by a linear model containing latent independent components and perturbed by an additive gaussian noise. The density of observed variables has been estimated by a kernel estimate by Amato et al. [8]. A linear latent variable model where observed variables depend linearly on unobservable latent variables has been analyzed in [9]. Under normality assumption the covariance structure of the model is estimated by maximum likelihood and its asymptotic normality is established. For ordered categorical data the latent variable model has been investigated by Breslaw and McIntosh [10] and by Gebregziabher and DeSantis [11] for missing categorical data. It has been applied to finance by Bai and Ng [12]. A generalized linear latent variable model (GLLVM) has been estimated using Laplace approximation by Bianconcini and Cagnone [13]. Similar model with semi-nonparametric specification of distribution of latent variables has been analyzed by Irincheeva, Cantoni and Genton [14]. Bartolucci, Pennoni and Francis [15] considered latent Markov model and estimated its parameters using EM algorithm and Bartolucci [16] applied it to detecting patterns of criminal activity.

A mixture of latent variables model was applied to clustering, classification and discriminant analysis; see [17]. Parsimonious Gaussian mixture models (PGMMs) are recently introduced model-based clustering techniques generalizing mixtures of factor analyzers model and are based on a latent Gaussian mixture model. McNicholas [18] used PGMM and Bayesian information criteria to perform model-based classification. A general latent variable model incorporating spatial correlation and shifted dependencies has been analyzed by Christensen and Amemiya [19]. Colombo et al. [20] applied latent variables to learning of high-dimensional acyclic graphs. In longitudinal data analysis one often encounters non-Gaussian data. Hall et al. [21] used a latent Gaussian process model for prediction by means of functional principal component analysis (PCA). The PCA approach has also been used to estimate latent variable models by Lynn and McCulloch [22]. In a model where the number of manifest variables is the same for all latent variables, and where this number and the number of observations of each of them increase, Bai and Ng [23] estimate the number of latent variables using an asymptotic principal component analysis.

The previous works on regression estimation in the context of latent variables were confined to parametric models, often formulated with so-called structural equations models; for surveys, see, e.g., Skrondal and Rabe-Hesketh [2] or Schumacker and Marcoulides [24]. In [25] a high-dimensional linear regression problem is considered, where a low dimensional latent variable model determines the response variable. Principal component analysis is used to estimate the underlying latent variables, and it is assumed that all variables have a Gaussian distribution. A generalization of Gaussian latent variable models to the case that the manifest variables are indirect observations of normal underlying variables can be done via generalized linear latent variable models; see, e.g., [26].

Our results generalize previously known results in so far that we do not need to impose any parametric structure on the regression function considered and that we do not restrict the class of error distributions occurring in the model. Our estimation of the common factor model is related to errors-in-variables models. In fact our estimation principle is based on generalization of the uniqueness result for such models presented in [27].

Nonparametric regression estimation has been studied in the literature for a long time. The most popular estimates for random design regression include kernel regression estimate (see, e.g., [28–33]), partitioning regression estimate (see, e.g., [34,35]), nearest neighbor regression estimate (see, e.g., [36–39]), least squares estimates (see, e.g., [40]) or smoothing spline estimates (see, e.g., [41]). The main theoretical results are summarized in the monograph by Györfi et al. [42]. To the best of the authors' knowledge, the application of nonparametric regression in the context of latent variables is new.

1.2. Notation

Throughout this paper we use the following notation: the sets of integers, rational numbers and real numbers are denoted by $\mathbb{N}$, $\mathbb{Q}$ and $\mathbb{R}$, respectively. For $k \in \mathbb{N}$ and subsets $B_1, \ldots, B_k$ of $\mathbb{R}^d$ we write

$$\prod_{i=1}^{k} B_i = \{(x_1, \ldots, x_k) : \forall_{i\in\{1,\ldots,k\}}\, x_i \in B_i\}$$

for the Cartesian product of the sets. $\mathbf{1}_B$ is the indicator of the set B. If X is an $\mathbb{R}^d$-valued random variable then $\varphi_X(u) = \mathbf{E}(e^{iu^\top X})$ is its characteristic function. If Z is an $\mathbb{R}$-valued random variable and $p \geq 1$ then we say that Z is in L_p if $\mathbf{E}(|Z|^p) < \infty$. For $f : D \to \mathbb{R}$ we write

$$x = \arg\min_{z\in D} f(z)$$

provided that $x \in D$ and $f(x) = \min_{z\in D} f(z)$.

1.3. Outline

The estimate of the common factor analysis model is described in Section 2. In Section 3 we use techniques of nonparametric regression to analyze the relationship between the latent variables. Section 4 illustrates the method by applying it to simulated data. The proofs are given in Section 5.

2. Estimation of a common factor analysis model

In the sequel X and Y are $\mathbb{R}^{d_X}$- and $\mathbb{R}^{d_Y}$-valued observable random variables (manifest variables). In order to analyze the relation between X and Y we assume that they depend linearly on some hidden and unobservable variables Z_1 and Z_2, where Z_1 and Z_2 are d_{Z_1}- and d_{Z_2}-dimensional random vectors, respectively. Here we assume $d_{Z_1} < d_X$ and $d_{Z_2} < d_Y$. More precisely we assume that X and Y satisfy the following common factor analysis models:

$$X = AZ_1 + \epsilon \tag{1}$$

and

$$Y = BZ_2 + \delta, \tag{2}$$

where A and B are $d_X \times d_{Z_1}$ and $d_Y \times d_{Z_2}$-dimensional matrices, respectively, and ϵ and δ are d_X- and d_Y-dimensional random vectors where all components are independent and have mean zero; furthermore we assume that (Z_1, Z_2), ϵ and δ are mutually independent. Given a sample $\mathcal{D}_n = \{(X_1, Y_1), \ldots, (X_n, Y_n)\}$ of independent and identically distributed copies of (X, Y), we want to estimate A, B and the corresponding values of the latent variables $Z_{1,i}$ and $Z_{2,i}$ corresponding to X_i and $Y_i (i = 1, \ldots, n)$. In the next section we will apply nonparametric regression to the estimated sample $\left\{(\hat{z}_{1,1}, \hat{z}_{2,1}), \ldots, (\hat{z}_{1,n}, \hat{z}_{2,n})\right\}$ of (Z_1, Z_2) in order to analyze the relation between Z_1 and Z_2.

In this section we describe how to estimate the common factor analysis model described by (1) and (2). Here we assume that some a priori information on the structure of the matrices is given. More precisely, we assume a so-called simple structure in terms of a single cause of variation (i.e., a single latent variable) for each manifest variables. The simple structure idea was coined by Thurstone [43] as a principle of factor rotation to improve the interpretability of latent variables as distinct attributes obtained from factor analytic procedures (for early discussions see, e.g., [44,45]).

In the meantime, the simple structure assumption plays an important (standard) role in the behavioral and social sciences, see, e.g., [46], which typically assume/intend homogeneous, uni-dimensional measures of latent variables (e.g., personality traits, competencies, attributes etc.). In other words, each of the components of the manifest variables is influenced by at most one of the components of the latent variables. Although this assumption is a common aim of test construction in the behavioral and social sciences, we have to emphasize that it is also a limitation of the proposed procedure.

The simple structure assumption can be relaxed to a certain extent by other semi-parametric or parametric latent variable modeling procedures (e.g., the Structural Equation Mixture Modeling approach, SEMM, [47]; the Latent Moderated Structural Equations approach, LMS, [48]; see also below). However, given the simple structure assumption, each row of A and B (as described by (1) and (2)) contains at most one nonzero entry. By rescaling the columns of the matrices and the latent variables we can assume furthermore that one of the entries in each column is 1 (which enables us to show that the model is uniquely defined, see Lemma 1). If this is true we can rewrite our model by (3) below, where we assume that $\ell_1, \ldots, \ell_{d_{Z_1}}, k_1, \ldots, k_{d_{Z_2}} \geq 3$.

Here $Z_1^{(1)}, \ldots, Z_1^{(d_{Z_1})}$ are the components of Z_1 and ℓ_i is the number of components of X influenced by $Z_1^{(i)}$. Similarly, $Z_2^{(1)}, \ldots, Z_2^{(d_{Z_2})}$ are the components of Z_2 and k_j is the number of components of Y influenced by $Z_2^{(j)}$. Furthermore we set

$$O = \left(X_{1,1}, \ldots, X_{1,\ell_1}, \ldots, X_{d_{Z_1},1}, \ldots, X_{d_{Z_1},\ell_{d_{Z_1}}}, Y_{1,1}, \ldots, Y_{1,k_1}, \ldots, Y_{d_{Z_2},1}, \ldots, Y_{d_{Z_2},k_{d_{Z_2}}}\right),$$

$$\begin{aligned} AZ = \big(&1Z_1^{(1)}, a_{1,2}Z_1^{(1)}, \ldots, a_{1,\ell_1}Z_1^{(1)}, \ldots, 1Z_1^{(d_{Z_1})}, a_{d_{Z_1},2}Z_1^{(d_{Z_1})}, \ldots, \\ &a_{d_{Z_1},\ell_{d_{Z_1}}}Z_1^{(d_{Z_1})}, 1Z_2^{(1)}, b_{1,2}Z_2^{(1)}, \ldots, b_{1,k_1}Z_2^{(1)}, \ldots, 1Z_2^{(d_{Z_2})}, b_{d_{Z_2},2}Z_2^{(d_{Z_2})}, \ldots, b_{d_{Z_2},k_{d_{Z_2}}}Z_2^{(d_{Z_2})}\big), \end{aligned}$$

$$E = (\epsilon_{1,1}, \ldots, \epsilon_{\ell_1,1}, \ldots, \epsilon_{1,d_{Z_1}}, \ldots, \epsilon_{\ell_{d_{Z_1}},d_{Z_1}}, \delta_{1,1}, \ldots, \delta_{k_1,1}, \ldots, \delta_{1,d_{Z_2}}, \ldots, \delta_{k_{d_{Z_2}},d_{Z_2}}).$$

Then our model can be rewritten as follows:

$$O^\top = AZ^\top + E^\top. \tag{3}$$

In order to simplify the notation we assume throughout that $d_{Z_1} = d_{Z_2} = 1$. Set

$$\begin{aligned} O_{simple} &= \left(X^{(1)}, \ldots, X^{(d)}, Y^{(1)}, \ldots, Y^{(\ell)}\right), \\ AZ_{simple} &= (1Z_1, a_2Z_1, \ldots, a_dZ_1, 1Z_2, b_2Z_2, \ldots, b_\ell Z_2)\,, \\ E_{simple} &= (\epsilon_1, \ldots, \epsilon_d, \delta_1, \ldots, \delta_\ell)\,. \end{aligned}$$

Consequently we can rewrite the model (1) and (2) in the following form:

$$O_{simple}^\top = AZ_{simple}^\top + E_{simple}^\top \tag{4}$$

where we assume that the coefficients are all nonzero, that $d, \ell \geq 3$, and that $Z_1, Z_2, \epsilon_1, \ldots, \epsilon_d, \delta_1, \ldots, \delta_\ell$ are real random variables with the property that $(Z_1, Z_2), \epsilon_1, \ldots, \epsilon_d, \delta_1, \ldots, \delta_\ell$ are independent and that satisfy $\mathbf{E}(\epsilon_j) = \mathbf{E}(\delta_k) = 0$.

Our first result shows that under the additional assumption that the characteristic function of

$$(X, Y) = (X^{(1)}, \ldots, X^{(d)}, Y^{(1)}, \ldots, Y^{(\ell)})$$

does not vanish at any point the distribution of (X, Y) determines uniquely the (joint) distribution of all other random variables occurring in the above model.

Lemma 1. *Assume that in the model (4) the random variables $X^{(1)}, \ldots, X^{(d)}, Y^{(1)}, \ldots, Y^{(\ell)}$ are in L_2, that $Z_1, Z_2, \epsilon_1, \ldots, \epsilon_d, \delta_1, \ldots, \delta_\ell$ are in L_1, that $(Z_1, Z_2), \epsilon_1, \ldots, \epsilon_d, \delta_1, \ldots, \delta_\ell$ are independent, that*

$$\mathbf{E}(\epsilon_1) = s = \mathbf{E}(\epsilon_d) = \mathbf{E}(\delta_1) = s = \mathbf{E}(\delta_\ell) = 0,$$

that $\mathbf{E}(Z_k^2) > 0$ $(k \in \{1, 2\})$ and that $a_2, \ldots, a_d, b_2, \ldots, b_\ell \in \mathbb{R}$ and $d, \ell \geq 3$ and $a_2 \neq 0$, $a_3 \neq 0$, $b_2 \neq 0$ and $b_3 \neq 0$. Assume furthermore, that the characteristic function of (X, Y) does not vanish at any point. Set

$$\begin{aligned} \widetilde{AZ}_{simple} &= (1\tilde{Z}_1, \tilde{a}_2\tilde{Z}_1, \ldots, \tilde{a}_d\tilde{Z}_1, 1\tilde{Z}_2, \tilde{b}_2\tilde{Z}_2, \ldots, \tilde{b}_\ell\tilde{Z}_2), \\ \tilde{E}_{simple} &= (\tilde{\epsilon}_1, \ldots, \tilde{\epsilon}_d, \tilde{\delta}_1, \ldots, \tilde{\delta}_\ell). \end{aligned}$$

If $\tilde{Z}_1, \tilde{Z}_2, \tilde{\epsilon}_1, \ldots, \tilde{\epsilon}_d, \tilde{\delta}_1, \ldots, \tilde{\delta}_\ell$ are in L_1, $\tilde{a}_2, \ldots, \tilde{a}_d, \tilde{b}_2, \ldots, \tilde{b}_\ell$ are in $\mathbb{R}$ and $\tilde{Z}_1, \ldots, \tilde{b}_\ell$ satisfy

$$O_{simple}^\top = \widetilde{AZ}_{simple}^\top + \tilde{E}_{simple}^\top$$

where the equality above holds in distribution,

$$\mathbf{E}(\tilde{\epsilon}_1) = s = \mathbf{E}(\tilde{\epsilon}_d) = \mathbf{E}(\tilde{\delta}_1) = s = \mathbf{E}(\tilde{\delta}_\ell) = 0$$

and $(\tilde{Z}_1, \tilde{Z}_2), \tilde{\epsilon}_1, \ldots, \tilde{\epsilon}_d, \tilde{\delta}_1, \ldots, \tilde{\delta}_\ell$ are independent, then $\tilde{a}_j = a_j$ $(j = 1, \ldots, d)$, $\tilde{b}_k = b_k$ $(k = 1, \ldots, \ell)$, $\mathbf{P}_{(\tilde{Z}_1,\tilde{Z}_2)} = \mathbf{P}_{(Z_1,Z_2)}$, $\mathbf{P}_{\tilde{\epsilon}_1} = \mathbf{P}_{\epsilon_1}, \ldots, \mathbf{P}_{\tilde{\epsilon}_d} = \mathbf{P}_{\epsilon_d}$ and $\mathbf{P}_{\tilde{\delta}_1} = \mathbf{P}_{\delta_1}, \ldots, \mathbf{P}_{\tilde{\delta}_\ell} = \mathbf{P}_{\delta_\ell}$.

Hence under the above assumptions $a_2, \ldots, a_d, b_2, \ldots, b_\ell$, and the distributions of $(Z_1, Z_2), \epsilon_1, \ldots, \epsilon_d, \delta_1, \ldots, \delta_\ell$ are uniquely determined by the distribution of (X, Y).

Remark 1. In case $d = 2$ and $\ell = 2$ the model (4) is not unique. For instance, if Z, ϵ_1 and ϵ_2 are independent normally distributed with mean zero, then the distribution of $(X_1, X_2)^\top = (Z + \epsilon_1, aZ + \epsilon_2)^\top$ does not uniquely determine the distribution of $Z, \epsilon_1, \epsilon_2$. For instance, take $a = 1, Z \sim \mathcal{N}(0, 1), \epsilon_1 \sim \mathcal{N}(0, 1), \epsilon_2 \sim \mathcal{N}(0, 4)$ or $a = 4, Z \sim \mathcal{N}(0, 1/4), \epsilon_1 \sim \mathcal{N}(0, 7/4), \epsilon_2 \sim \mathcal{N}(0, 1)$. By computing covariance matrices it is easy to see that in both cases the distributions of (X_1, X_2) are the same.

Remark 2. A generalization of the proof of Lemma 1 shows that if we assume the model (3) in case $d_{z_1} > 1$ or $d_{z_2} > 1$, then our independence assumption together with the assumption that the characteristic function does not vanish imply that the distribution of (X, Y) uniquely determines the joint distribution of all other variables occurring in the model and all coefficients $a_{i,\ell}$ and $b_{j,k}$.

In the sequel we want to estimate the above latent variable model from the independent and identically distributed observations $(X_1, Y_1), \ldots, (X_n, Y_n)$.

The crucial property which allows us to show that the above model is uniquely determined is independence of the random variables. In the sequel we use this property for estimation of the model by determining estimates of the values of the latent variables in such a way that the corresponding empirical distributions satisfy asymptotically this independence assumption.

We start with definition of the estimate of the above model by estimating the coefficients a_j and b_k. Here we use

$$a_2 = \frac{\mathbf{E}\{X^{(2)}X^{(3)}\}}{\mathbf{E}\{X^{(1)}X^{(3)}\}}, \qquad a_j = \frac{\mathbf{E}(X^{(2)}X^{(j)})}{\mathbf{E}(X^{(1)}X^{(2)})}$$
$$b_2 = \frac{\mathbf{E}\{Y^{(2)}Y^{(3)}\}}{\mathbf{E}\{Y^{(1)}Y^{(3)}\}}, \qquad b_k = \frac{\mathbf{E}(Y^{(2)}Y^{(k)})}{\mathbf{E}(Y^{(1)}Y^{(2)})}$$

for $j, k > 2$; see the proof of Lemma 1. We also set $\hat{a}_1 = \hat{b}_1 = 1$ and, for $j, k > 2$,

$$\hat{a}_2 = \frac{\sum_{i=1}^n X_i^{(2)}X_i^{(3)}}{\sum_{i=1}^n X_i^{(1)}X_i^{(3)}}, \qquad \hat{a}_j = \frac{\sum_{i=1}^n X_i^{(2)}X_i^{(j)}}{\sum_{i=1}^n X_i^{(1)}X_i^{(2)}}$$

and

$$\hat{b}_2 = \frac{\sum_{j=1}^n Y_j^{(2)}Y_j^{(3)}}{\sum_{j=1}^n Y_j^{(1)}Y_j^{(3)}}, \qquad \hat{b}_k = \frac{\sum_{j=1}^n Y_j^{(2)}Y_j^{(k)}}{\sum_{j=1}^n Y_j^{(1)}Y_j^{(2)}}.$$

Next we try to determine estimates $(\hat{z}_{1,i}, \hat{z}_{2,i})$ of $(Z_{1,i}, Z_{2,i})$ for $i = 1, \ldots, n$. As soon as such estimates are available, we also have estimates of the values of $\epsilon_j = X^{(j)} - a_j Z_1$ and $\delta_k = Y^{(k)} - b_k Z_2$, namely

$$\hat{\epsilon}_{j,i} = X_i^{(j)} - \hat{a}_j \hat{z}_{1,i} \quad \text{and} \quad \hat{\delta}_{k,i} = Y_i^{(k)} - \hat{b}_k \hat{z}_{2,i}$$

$(i = 1, \ldots, n)$, so we have available an estimated sample of the joint distribution of $((Z_1, Z_2), \epsilon_1, \ldots, \epsilon_d, \delta_1, \ldots, \delta_\ell)$.

The basic idea is to consider the empirical distribution μ_n belonging to this estimated sample and to determine the estimates of the values of the latent variables in such a way that this empirical distribution satisfies approximately the independence condition of Lemma 1 and $\mathbf{E}(\epsilon_j) = \mathbf{E}(\delta_k) = 0$ which ensure uniqueness of the latent variable model.

More precisely, for values $\kappa_1, \ldots, \kappa_n$ in $\mathbb{R}^p$ let μ_{n,κ_1^n} be the empirical distribution of $\kappa_1, \ldots, \kappa_n$, i.e., for $B \subseteq \mathbb{R}^p$, set

$$\mu_{n,\kappa_1^n}(B) = \frac{1}{n}\sum_{i=1}^n \mathbf{1}_B(\kappa_i).$$

Let $\hat{\mu}_n^{(\hat{z}_1,\hat{z}_2)_1^n}$ be the empirical distribution corresponding to the pseudo-sample defined, for each $i \in \{1, \ldots, n\}$, by

$$\begin{aligned}&((\hat{z}_{1,i}, \hat{z}_{2,i}), \hat{\epsilon}_{1,i}, \ldots, \hat{\epsilon}_{d,i}, \hat{\delta}_{1,i}, \ldots, \hat{\delta}_{\ell,i})\\ &\quad = ((\hat{z}_{1,i}, \hat{z}_{2,i}), X_i^{(1)} - \hat{a}_1\hat{z}_{1,i}, \ldots, X_i^{(d)} - \hat{a}_d\hat{z}_{1,i}, Y_i^{(1)} - \hat{b}_1\hat{z}_{2,i}, \ldots, Y_i^{(\ell)} - \hat{b}_\ell\hat{z}_{2,i})\end{aligned}$$

i.e.,

$$\hat{\mu}_n^{(\hat{z}_1,\hat{z}_2)_1^n} = \mu_{n,((\hat{z}_1,\hat{z}_2),\hat{\epsilon}_1,\ldots,\hat{\epsilon}_d,\hat{\delta}_1,\ldots,\hat{\delta}_\ell)_1^n}.$$

The distribution μ of $((Z_1, Z_2), \epsilon_1, \ldots, \epsilon_d, \delta_1, \ldots, \delta_\ell)$ satisfies

$$\mu\left(\prod_{i=1}^{1+d+\ell} B_i\right) = \mu\left(B_1 \times \prod_{j=2}^{1+d+\ell} \mathbb{R}\right) \prod_{i=2}^{1+d+\ell} \mu\left(\mathbb{R}^2 \times \prod_{j=2}^{i-1} \mathbb{R} \times B_i \times \prod_{j=i+1}^{1+d+\ell} \mathbb{R}\right)$$

for any $B_1 \in \mathcal{B}_2, B_2 \in \mathcal{B}, \ldots, B_{1+d+\ell} \in \mathcal{B}$ because of the independence assumption, where $\mathcal{B}$ and $\mathcal{B}_2$ denote the Borel σ–field in $\mathbb{R}$ and in $\mathbb{R}^2$, respectively. It follows from the Carathéodory's extension theorem that if this relation holds for all intervals of the form $(-\infty, x]$, then μ has independent components. We choose our estimated values such that this is approximately true for the empirical distribution $\hat{\mu}_n^{(\hat{z}_1,\hat{z}_2)_1^n}$. More precisely, we choose $\hat{\mu}_n^{(\hat{z}_1,\hat{z}_2)_1^n}$ such that

$$\hat{\mu}_n^{(\hat{z}_1,\hat{z}_2)_1^n}\left(\prod_{i=1}^{1+d+\ell} B_i\right) - \hat{\mu}_n^{(\hat{z}_1,\hat{z}_2)_1^n}\left(B_1 \times \prod_{j=2}^{1+d+\ell} \mathbb{R}\right) \prod_{i=2}^{1+d+\ell} \hat{\mu}_n^{(\hat{z}_1,\hat{z}_2)_1^n}\left(\mathbb{R}^2 \times \prod_{j=2}^{i-1} \mathbb{R} \times B_i \times \prod_{j=i+1}^{1+d+\ell} \mathbb{R}\right) \approx 0$$

holds for suitably chosen sets $B_1, \ldots, B_{1+d+\ell} \in \mathcal{B}$. In order to be able to compute the estimate, we use here a sigmoidal approximation of the indicator function of an interval.

More precisely, we choose a continuous sigmoidal function $\sigma : \mathbb{R} \to \mathbb{R}$, i.e., a continuous monotone function $\sigma : \mathbb{R} \to [0, 1]$ satisfying $\sigma(x) \to 0$ as $x \to -\infty$ and $\sigma(x) \to 1$ as $x \to \infty$, probability weights $(p_r)_{r\in\mathbb{N}}$, $\alpha_{r,1}, \alpha_{r,2}, \beta_{r,j}, \gamma_{r,k} \in \mathbb{Q}$ such that

$$\mathbb{Q}^{2+d+\ell} = \{(\alpha_{r,1}, \alpha_{r,2}, \beta_{r,1}, \ldots, \beta_{r,d}, \gamma_{r,1}, \ldots, \gamma_{r,\ell}) : r \in \mathbb{N}\}$$

and $N_n \in \mathbb{N}$ satisfying $N_n \to \infty (n \to \infty)$, and define, for each $i \in \{1, \ldots, n\}$, $(\hat{z}_{1,i}, \hat{z}_{2,i})$ as the value which minimizes

$$\begin{aligned} T_n = \sum_{r=1}^{N_n} \Bigg| & \frac{1}{n}\sum_{i=1}^{n} \sigma\{-n(z_{1,i} - \alpha_{r,1})\}\sigma\{-n(z_{2,i} - \alpha_{r,2})\} \prod_{j=1}^{d} \sigma\{-n(\epsilon_{j,i} - \beta_{r,j})\} \prod_{k=1}^{\ell} \sigma\{-n(\delta_{k,i} - \gamma_{r,k})\} \\ & - \frac{1}{n}\sum_{i=1}^{n} \sigma\{-n(z_{1,i} - \alpha_{r,1})\}\sigma\{-n(z_{2,i} - \alpha_{r,2})\} \prod_{j=1}^{d} \frac{1}{n}\sum_{i=1}^{n} \sigma\{-n(\epsilon_{j,i} - \beta_{r,j})\} \prod_{k=1}^{\ell} \frac{1}{n}\sum_{i=1}^{n} \sigma\{-n(\delta_{k,i} - \gamma_{r,k})\} \Bigg|^2 p_r \\ & + \sum_{j=1}^{d} \left(\frac{1}{n}\sum_{i=1}^{n} \epsilon_{j,i}\right)^2 + \sum_{k=1}^{\ell} \left(\frac{1}{n}\sum_{i=1}^{n} \delta_{k,i}\right)^2 \end{aligned}$$

with respect to $(z_{1,i}, z_{2,i})$ subject to the constraints

$$\frac{1}{n}\sum_{i=1}^{n} z_{1,i}^2 \le 1 + \frac{1}{n}\sum_{i=1}^{n} (X_i^{(1)})^2 \quad \text{and} \quad \frac{1}{n}\sum_{i=1}^{n} z_{2,i}^2 \le 1 + \frac{1}{n}\sum_{i=1}^{n} (Y_i^{(1)})^2, \tag{5}$$

where

$$\epsilon_{j,i} = X_i^{(j)} - \hat{a}_j z_{1,i} \quad \text{and} \quad \delta_{k,i} = Y_i^{(k)} - \hat{b}_k z_{2,i}.$$

Our main result is the following theorem.

Theorem 1. *Assume that the assumptions of Lemma 1 are satisfied, and let the estimate $\hat{\mu}_n^{(\hat{z}_1,\hat{z}_2)_1^n}$ of the distribution μ of $((Z_1, Z_2), \epsilon_1, \ldots, \epsilon_d, \delta_1, \ldots, \delta_\ell)$ be defined as above. Then, with probability 1,*

$$\hat{\mu}_n^{(\hat{z}_1,\hat{z}_2)_1^n} \to \mu \quad \textit{weakly},$$

i.e., as $n \to \infty$, $\hat{\mu}_n^{(\hat{z}_1,\hat{z}_2)_1^n}(A) \to \mu(A)$ for all sets A such that the boundary ∂A satisfies $\mu(\partial A) = 0$.

Remark 3. It is straightforward to extend our estimate to the case of model (3) with $d_{Z_1} > 1$ or $d_{Z_2} > 1$. In order to estimate the coefficients $a_{r,s}$ and $b_{r,s}$ in (3), one replaces in definition of $\hat{a}_j$ and $\hat{b}_k$ the random variables $X^{(1)}, X^{(2)}, X^{(3)}$ and $Y^{(1)}, Y^{(2)}, Y^{(3)}$ by $X_{r,1}, X_{r,2}, X_{r,3}$ and $Y_{r,1}, Y_{r,2}, Y_{r,3}$, respectively. In order to estimate the values of latent variables one just needs to replace the empirical distribution of $((\hat{z}_{1,i}, \hat{z}_{2,i}), \hat{\epsilon}_{1,i}, \ldots, \hat{\epsilon}_{d,i}, \hat{\delta}_{1,i}, \ldots, \hat{\delta}_{\ell,i})$ by the empirical distribution of the vector of all latent variables and all estimated error terms in model (3) and adjust the definition of T_n.

Remark 4. In our definition of the estimate we minimize T_n subject to constraint (5). It follows from the proof of Theorem 1 that we can impose even more restrictions in the above minimization problems, as long as the values of the latent variables satisfy them with probability 1 for large n. For instance, in the next section we will assume $\mathbf{E}(|Y^{(1)}|^4) < \infty$. Since Z_2 and δ_1 are independent, $Y^{(1)} = Z_2 + \delta_1$ and $\mathbf{E}(\delta_1) = 0$, this implies

$$\mathbf{E}\{|Y^{(1)} - \mathbf{E}(Y^{(1)})|^4\} = \mathbf{E}\{|Z_2 - \mathbf{E}(Z_2) + \delta_1|^4\} \ge \mathbf{E}\{|Z_2 - \mathbf{E}(Z_2)|^4\},$$

hence

$$\begin{aligned} \mathbf{E}(Z_2^4) &\le 2^4 \mathbf{E}\{(Z_2 - \mathbf{E}Z_2)^4\} + 2^4 |\mathbf{E}(Y^{(1)})|^4 \\ &\le 256\mathbf{E}(|Y^{(1)}|^4) + 272|\mathbf{E}(Y^{(1)})|^4. \end{aligned}$$

Consequently, if we impose in this case the additional constraint

$$\frac{1}{n}\sum_{i=1}^{n} \hat{z}_{2,i}^4 \le 1 + 256\frac{1}{n}\sum_{i=1}^{n} (Y_i^{(1)})^4 + 272\left(\frac{1}{n}\sum_{i=1}^{n} Y_i^{(1)}\right)^4 \tag{6}$$

in the above minimization problem, then the assertion of Theorem 1 still holds.

3. Estimation of the regression function corresponding to latent variables

In this section we estimate the regression function corresponding to the latent variables Z_1 and Z_2 in model (4), i.e., we estimate

$$m : \mathbb{R} \to \mathbb{R}, \qquad m(x) = \mathbf{E}(Z_2|Z_1 = x),$$

from the data $\mathcal{D}_n = \{(X_1, Y_1), \ldots, (X_n, Y_n)\}$. The basic idea is to use the data as in Section 2 to construct the sample $(\hat{z}_{1,1}, \hat{z}_{1,2}), \ldots, (\hat{z}_{n,1}, \hat{z}_{n,2})$ of (Z_1, Z_2) and to apply a regression estimate to this data.

By Theorem 1 we know that in case that we assume that all occurring random variables are bounded

$$\frac{1}{n}\sum_{i=1}^{n} |\hat{z}_{i,2} - f(\hat{z}_{i,1})|^2 - \mathbf{E}|Z_2 - f(Z_1)|^2 = \int |z_2 - f(z_1)|^2 d\hat{\mu}_n^{(\hat{z}_1,\hat{z}_2)} - \int |z_2 - f(z_1)|^2 d\mu \to 0$$

a.s. for all bounded and continuous functions $f : \mathbb{R} \to \mathbb{R}$. We will see in the proof of Theorem 2 that when we impose the additional constraint (6) in the definition of our estimate, then this result also holds for unbounded random variables provided that $\mathbf{E}(|Y^{(1)}|^4) < \infty$.

Since

$$\mathbf{E}\{|Z_2 - m(Z_1)|^2\} = \min_{f:\mathbb{R}\to\mathbb{R}} \mathbf{E}\{|Z_2 - f(Z_1)|^2\}$$

(see, e.g., Section 1.1 in Györfi et al. [42]) this motivates to estimate the regression function m by the well-known least squares estimate

$$m_n() = \arg\min_{f\in\mathcal{F}_n} \frac{1}{n}\sum_{i=1}^{n} |\hat{z}_{i,2} - f(\hat{z}_{i,1})|^2, \tag{7}$$

where $\mathcal{F}_n$ is a suitable defined set of functions consisting of continuous and bounded functions $f : \mathbb{R} \to \mathbb{R}$ depending on the sample size n. For notational simplicity we assume here and in the sequel that the minimum above exists. Our main result is the following theorem.

Theorem 2. *Assume that in the model (4) the random variables $Z_1, Z_2, \epsilon_1, \ldots, \epsilon_d, \delta_1, \ldots, \delta_\ell$ are in L_1, that $(Z_1, Z_2), \epsilon_1, \ldots, \epsilon_d, \delta_1, \ldots, \delta_\ell$ are independent, that*

$$\mathbf{E}(\epsilon_1) = s = \mathbf{E}(\epsilon_d) = \mathbf{E}(\delta_1) = s = \mathbf{E}(\delta_\ell) = 0,$$

that $\mathbf{E}(Z_k^2) > 0$ ($k \in \{1, 2\}$) and that $a_2, \ldots, a_d, b_2, \ldots, b_\ell \in \mathbb{R}$ and $d, \ell \geq 3$ and $a_2 \neq 0$, $a_3 \neq 0$, $b_2 \neq 0$ and $b_3 \neq 0$. Assume furthermore that the characteristic function of (X, Y) does not vanish at any point, that $X^{(1)}, \ldots, X^{(d)}, Y^{(1)}, \ldots, Y^{(\ell)}$ are in L_2 and that $\mathbf{E}(|Y^{(1)}|^4) < \infty$.

Let $\mathcal{F}_n$ be sets of functions $f : \mathbb{R}^d \to \mathbb{R}$ which are bounded by some constant $L > 0$ and assume that

$$\bigcup_{n=1}^{\infty} \mathcal{F}_n \quad \text{is an equicontinuous set of functions.} \tag{8}$$

Let the least squares estimate m_n be defined as above, where we impose the condition (6) as an additional constraint in the minimization problem. If, as $n \to \infty$,

$$\inf_{f\in\mathcal{F}_n} \int |f(x) - m(x)|^2 \mathbf{P}_{Z_1}(dx) \to 0 \tag{9}$$

then

$$\int |m_n(x) - m(x)|^2 \mathbf{P}_{Z_1}(dx) \to 0 \quad a.s.$$

In the sequel we choose $\mathcal{F}_n$ as a suitably defined space of polynomial splines and show that in the case of bounded and Lipschitz continuous regression functions the corresponding least squares estimate (7) is strongly consistent.

Let $M \in \mathbb{N}$ be arbitrary. For $j \in \mathbb{Z}$ and $K \in \mathbb{N}$ let $B_{j,M}^K : \mathbb{R} \to \mathbb{R}$ be the B-spline with degree M, knot sequence $\{i/K : i \in \mathbb{Z}\}$ and support $[j/K, (j + M + 1)/K]$; see, e.g., [49,50] or Chapter 14 in Györfi et al. [42]. One well-known property of B-splines is that they are nonnegative and sum up to one (see de Boor [49, pp. 109,110]). Furthermore,

$$\left\{ \sum_{i=-M}^{K-1} a_i B_{i,M}^K : a_i \in \mathbb{R} \right\}$$

is on [0, 1] equal to the set of all piecewise polynomials of degree M with respect to a partition of [0, 1] consisting of K equidistant intervals, which are $(M-1)$–times continuously differentiable on [0, 1]. For $K_n \in \mathbb{N}$, $c_1 > 0$ and $c_2 > 0$ set

$$\mathcal{F}_n = \left\{ \sum_{j=-M}^{K_n-1} a_j B_{j,M}^{K_n} : |a_j - a_{j-1}| \leq \frac{c_1}{K_n} \text{ and } |a_j| \leq c_2 \quad (j \in \mathbb{Z}) \right\} \tag{10}$$

and define the estimate m_n by (7). Then the following result holds.

Corollary 1. *Assume that the assumptions of Theorem 1 are valid, and, in addition, that $m(x) = \mathbf{E}(Z_2|Z_1 = x)$ is Lipschitz continuous and bounded in absolute value. Assume furthermore that $Z_1 \in [0, 1]$ a.s. and that we enforce in the definition of the estimate in Section 2 $\hat{z}_{i,1} \in [0, 1]$ for all $i \in \{1, \ldots, n\}$. Let the least squares estimate m_n be defined as in Theorem 2 for some $K_n > 0$ satisfying $K_n \to \infty$ as $n \to \infty$. Then for c_1 and c_2 sufficiently large we have*

$$\int |m_n(x) - m(x)|^2 \mathbf{P}_{Z_1}(dx) \to 0 \quad a.s.$$

Proof. The functions in $\mathcal{F}_n$ are Lipschitz continuous with Lipschitz constant c_1 (see, e.g., Lemma 14.6 in Györfi et al. [42]), hence $\cup_{n=1}^{\infty} \mathcal{F}_n$ is equicontinuous. Furthermore, they are all bounded in absolute value by L (see, e.g., Lemmas 14.2 and 14.4 in Györfi et al. [42]). The result follows from Theorem 2, given that as $n \to \infty$,

$$\inf_{f \in \mathcal{F}_n} \int |f(x) - m(x)|^2 \mathbf{P}_{Z_1}(dx) \leq \inf_{f \in \mathcal{F}_n} \sup_{x \in [0,1]} |f(x) - m(x)|^2 \to 0,$$

which follows because of m Lipschitz continuous and c_1 and c_2 sufficiently large from $K_n \to \infty$ as $n \to \infty$; see, e.g., Györfi et al. [42, p. 271]. □

Remark 5. Any application of the above estimate requires a data-dependent choice of all parameters of the functions space, in particular of the bounds on the coefficients and the differences of the coefficients. One way of doing this is to use splitting of the sample. It is an open problem whether in this case the above consistency result still holds or (in case that it is not valid) whether there exists another method for a data-dependent choice of the parameters leading to consistent estimates.

Remark 6. If we estimate in Remark 3 a latent variable model where $d_{Z_1} > 1$, we can use the tensor product splines (see, e.g., Chapter 15 of Györfi et al. [42]) to estimate the multivariate regression function corresponding to Z_1 and Z_2 in an analogous way as before. This approach typically suffers from the curse of dimensionality, but following Stone [51,52] we can impose additional constraints on the structure of the regression function in order to get good results even for large values of d_{Z_1}.

4. Application to simulated data

4.1. Aim of simulation studies and selected approaches

This section describes three simulation studies that compare the proposed nonparametric approach with two alternative approaches and show the capability of a multidimensional estimation, respectively. The aim of the simulations studies is to examine the robustness of the approaches in the context of varying regression functions and varying non-normal distributions when nonlinear relations of the latent variables are approximated.

As one alternative approach, the parametric Latent Moderated Structural Equations approach (LMS) was applied as described by Klein and Moosbrugger [48]. The LMS approach is the standard procedure in the social and behavioral sciences when models involving products of latent variables (i.e., polynomials of second degree) are estimated. LMS is capable of estimating simple parametric relations between latent variables involving latent product terms (e.g., $Z_2 = \alpha + \gamma_1 Z_1 + \gamma_2 Z_1^2 + \eta$). Assuming normally distributed predictor variables (here Z_1), LMS provides maximum likelihood parameter estimates of the model. The key idea is to approximate the likelihood of the non-normally distributed indicator vector (which is always non-normally distributed due to the latent product term) by a finite mixture of conditionally normal distributions. Thereby, the latent variable (here Z_1), which is involved in the product term, is used as the conditioning variable. Using the expectation–maximization algorithm (EM; [53]) the likelihood is maximized; for technical details see [54,48]. When normally distributed latent predictor variables are given and when the parametric relationship with simple product terms is correctly specified, LMS is known to produce unbiased, consistent, as well as efficient parameter estimates [55,48]. Especially in the second simulation study, in which simple polynomial quadratic effects were given, LMS served as a comparative method. LMS is implemented as a standard routine in the commercial `Mplus` software (`XWITH` command; [56]).

As the second alternative approach, the semi-parametric Structural Equation Mixture Modeling approach (SEMM) described in [47,57,58] was selected. The SEMM approach uses mixtures of linear structural equation models (e.g., $Z_{2g} = \gamma_{1g} Z_{1g} + \eta_g$ for a component g) to approximate the unknown non-linear relationship of the latent variables. While parametric non-linear latent variable approaches (such as LMS) specify the functional form of their relationship a priori, the SEMM

approach does not require assumptions about the functional form. Only linearity of the latent variables within each component g is assumed. The total number of components (G) is not set a priori. The SEMM approach starts with estimating only one mixture component and iteratively increases the number of components for the next model after a solution was obtained (again, using the EM algorithm). Models with varying numbers of linear mixture components are compared using information criteria (AIC or BIC). Altogether, the SEMM approach does not require the assumption of normally distributed latent predictor and outcome variables and disturbances. It allows for a simultaneous flexible approximation of non-linear latent variable relationships (with one latent predictor and one latent outcome variable) and non-normal distributions of the latent variables by aggregating over the (linear) mixture components. Due to this aggregation of the mixture components, the SEMM approach has the limitation that a separate approximation of non-linearity and non-normality is not possible; for a discussion see [47]. By now, there has been one simulation study by [59] showing that the SEMM approach has a better performance than the LMS approach, when the functional form of the regression function is not a typical polynomial relationship with simple product terms of the latent variables. When all assumptions of the LMS approach (see above) are met, the LMS approach provides unbiased, efficient, and consistent estimates.

In the end the properties of the discussed approaches can be summarized as follows: LMS is a parametric approach which approximates the non-normal likelihood of the indicator vector using a finite mixture procedure. It assumes product terms of the latent variables in a polynomial regression function of the second order and normally distributed latent predictor variables. SEMM is a flexible semi-parametric mixture structural equation modeling approach. The mixtures of linear structural equation models are used to approximate simultaneously the non-linearity and non-normality. Although the proposed SEMM approach was able to handle only one-dimensional predictor variables, its extension to the multivariate case is straightforward (and we have extended the SEMM approach in our third simulation study in this way). The nonparametric approach is not a mixture approach to non-normality and non-linearity of the latent variables. As a two-step procedure, estimation of the (non-) normal latent variables and estimation of the latent non-linear relationship are separated. As a result, fewer assumptions about distributions and functional forms are imposed by the proposed nonparametric approach. However, one limitation of the proposed nonparametric approach, i.e., the assumption of a simple structure of the loading coefficients in the measurement models (see (1) and (2)), can be relaxed in the LMS and the SEMM approach. They are both to some extent capable of cross-loadings on their predictor sides. In other words, the coefficient matrix A cannot be estimated completely free (with all elements being free parameters) which would result in local identification issues. However, the capability of specifying cross-loadings offers a flexibility in the case of heterogeneous measures for the LMS and the (extended) SEMM approaches.

The SEMM approach with G components was implemented in the `Mplus` software [56] as described by Bauer [47] and Pek et al. [58]. According to Bauer [47] and Bauer et al. [57], for each replication per condition 500 random starting values were generated and maximum likelihood estimates were obtained. The solution with the highest likelihood was selected. The smoothed regression function values were obtained from the estimates. For each replication, the number of components G was determined by fitting models with an increasing number of components. The AIC and BIC information criteria were used to select the number of components. For each criterion separately, the number of components was increased until a minimum was achieved. Bauer et al. [57] stated that the AIC criterion might be better for indirect applications of SEMM for model selection (which generally favors more classes when components are used as an approximation device). The BIC could be used for the detection of the true number of components in direct applications.

The proposed nonparametric approach was implemented in MATLAB.[1] The first step after estimating $\hat{a}$ and $\hat{b}$ was to calculate $\hat{z}_1$ and $\hat{z}_2$ by minimizing T_n using an interior-point method; see, e.g., [60]. To find a starting point, a constant linear regression was applied to $(\hat{a}_1, X_i^1), \ldots, (\hat{a}_d, X_i^d)$ and $(\hat{b}_1, Y_i^1), \ldots, (\hat{b}_d, Y_i^\ell)$, respectively ($i = 1, \ldots, n$). The number of summands N_n was taken as $n^{1/3}$ and $p_r = 1/N_n$. The probability weights α_r, $\beta_{r,d}$, and $\gamma_{r,\ell}$ ($r = 1, \ldots, N_n$) were randomly generated, each having standard normal distribution. Since the given problem is nonconvex it is possible that the interior-point method does not find the global minimum, but rather a local minimum. To handle this problem the optimization can be carried out with several distinct starting points. They could simply be chosen by setting

$$\hat{z}_{1,i} = \left\{ n + \sum_{i=1}^{n} (X_i^{(1)})^2 \right\}^{1/2} \quad \text{and} \quad \hat{z}_{2,i} = \left\{ n + \sum_{i=1}^{n} (Y_i^{(1)})^2 \right\}^{1/2}$$

for arbitrary $i \in \{1, \ldots, n\}$. If it does not fit the additional constraint (6) one could subtract the necessary value to fulfill it. These choices however could greatly increase the duration of the optimization process. If the values found during optimization do not differ (within the range of 0.2%) it is plausible to assume that the given solution is the global optimum and if they differ the solution with the lowest value is chosen. As soon as the minimum of T_n was reached, the new probability weights were randomly chosen and the minimum was recomputed. The process was repeated several times. As long as new random parameters did not significantly ameliorate the minimum, the first choice was assumed to be reliable. In the next step, nonparametric regression based on B-splines was used to estimate the relation between the latent variables.

[1] The code can be downloaded from https://github.com/tifasch.

4.2. Design of the simulation studies and data generation

4.2.1. Simulation study 1

In the first simulation study, the latent predictor variable Z_1 was uniformly distributed on [0, 1]. The relation between the latent variables (Z_1 and Z_2) was given by three different regression functions:

$$Z_2 = \sin(2\pi Z_1) + 0.75\,\eta \tag{11}$$

$$Z_2 = \frac{1}{5}\,e^{5Z_1} - 25\,Z_1^3 + 0.75\,\eta \tag{12}$$

$$Z_2 = \frac{1}{5Z_1 + 1} + \sin(5\,Z_1) + 0.75\,\eta, \tag{13}$$

with $\eta \sim \mathcal{N}(0, 1)$. The three regression functions are later referred to as sin2pi, exp, and sin functions, respectively.

Each latent variable was represented by three observed variables, which were generated according to the following measurement models:

$$X_i = a_i Z_1 + 0.15\epsilon_i \tag{14}$$

$$Y_i = b_i Z_2 + \delta_i \tag{15}$$

$i = 1, 2, 3$, where $a = (1, 1.3, 1.8)$ and $b = (1, 1.1, 1.7)$. Random variables ϵ_i and δ_i are independent and have standard normal distribution. The sample size was set equal to $N = 400$. For each of the three regression functions, 200 data sets were generated. The SEMM approach, the LMS approach, and the proposed nonparametric approach were applied to the data.

4.2.2. Simulation study 2

In the second simulation study, we examined the robustness of the approaches for varying degrees of non-normality of the latent predictor variable Z_1. Non-normality of the latent predictor variable Z_1 was induced using the Vale and Maurelli [61] method. Three conditions of univariate skewness and kurtosis were selected for the latent predictor variable (in line with the values used by Curran et al. [62]): (a) normality with skewness 0 and kurtosis 0, (b) moderate non-normality with skewness 2 and kurtosis 7, and (c) strong non-normality with skewness 3 and kurtosis 21.

The relation between the latent variables (Z_1 and Z_2) was given by the following regression function (see [57]):

$$Z_2 = 5 - .5\,Z_1^2 + \eta, \tag{16}$$

where $\mathrm{E}(Z_1) = 0$, $\mathrm{var}(Z_1) = 1$ and $\eta \sim \mathcal{N}(0, .5)$.

Each latent variable (Z_1 and Z_2) was represented by three observed variables, which were generated according to the following measurement models:

$$X_i = Z_1 + \epsilon_i \tag{17}$$

$$Y_i = Z_2 + \delta_i \tag{18}$$

$i = 1, 2, 3$. Again, ϵ_i and δ_i are independent standard normal random variables. The sample size was varied at three levels ($N = 250$, 500, and 1000). For each of the resulting nine conditions, 250 data sets (replications) were generated. The SEMM approach, the LMS approach, and the proposed nonparametric approach were applied to the data.

4.2.3. Simulation study 3

In the third simulation study, the latent predictor variable $Z_1 = (Z_{11}, Z_{12})^\top$ was chosen two-dimensional (with two independent latent components Z_{11} and Z_{12}). The two latent predictor variables Z_{11} and Z_{12} were either both uniformly distributed on [0, 1] or both standard normally distributed, to show that the nonparametric approach works in both cases.

The relation between the latent variables (Z_1 and Z_2) was given by two different regression functions:

$$Z_2 = Z_{11}\sin(Z_{11}^2) - Z_{12}\sin(Z_{12}^2) + 0.15\,\eta \tag{19}$$

$$Z_2 = \frac{4}{1 + 4(Z_{11} - 0.5)^2 + 4(Z_{12} - 0.5)^2} + 0.15\,\eta, \tag{20}$$

with $\eta \sim \mathcal{N}(0, 1)$. The two regression functions are later referred to as sin2dim and quad, respectively.

Each latent variable (Z_{11}, Z_{12} and Z_2) was represented by observed variables, which were generated according to the following measurement models:

$$X_{1,i} = a_{1,i} Z_{11} + 0.15\,\epsilon_{1,i} \tag{21}$$

$$X_{2,j} = a_{2,j} Z_{12} + 0.15\,\epsilon_{2,j} \tag{22}$$

$$Y_i = b_i Z_2 + 0.2\,\delta_i \tag{23}$$

Table 1
Results for the simulation study with varying regression functions and a uniform distribution of Z_1.

		sin2pi	exp	sin
Bias	N_PAR	**0.099**	**0.466**	**0.042**
	AIC	1.004	1.327	1.031
	BIC	0.593	1.327	0.647
	LMS	1.041	3.247	2.357
SD	N_PAR	**0.165**	0.249	**0.160**
	AIC	0.656	0.225	0.942
	BIC	1.076	**0.225**	1.002
	LMS	0.325	0.683	0.346
RMSE	N_PAR	**0.192**	**0.528**	**0.166**
	AIC	1.200	1.346	1.192
	BIC	1.229	1.346	1.396
	LMS	1.090	3.318	2.382

Note. N_PAR = nonparametric approach; AIC = SEMM approach with AIC; BIC = SEMM approach with BIC; LMS = Latent Moderated Structural Equations approach. sin2pi represents the sinusoidal function in Eq. (11). exp represents the exponential function in Eq. (12). sin represents the sinusoidal function in Eq. (13). The lowest values of bias, SD, and RMSE are bolded within each case (column).

$i = 1, 2, 3$ and $j = 1, 2, 3, 4$, where $a_1 = (1, 1.3, 1.8)$, $a_2 = (1, 1.5, 2, 1.3)$ and $b = (1, 1.1, 1.7)$. Random variables $\epsilon_{1,i}$, $\epsilon_{2,j}$ and δ_i were independent and had standard normal distributions.

The sample size was set equal to $N = 500$. For each of the two regression functions, 200 data sets were generated. The SEMM approach, the LMS approach, and the proposed nonparametric approach were applied to the data.

4.3. Results of the simulation studies

4.3.1. Simulation study 1

Table 1 presents the bias, standard deviation (SD) and root mean square error (RMSE) of the regression function estimates for the three different conditions. For an easier interpretation the lowest values for each condition (in columns) are bolded.

The results can be summarized as follows: First, the nonparametric approach showed a lower bias in all cases than the SEMM approach and the LMS approach. The SEMM approach performed better than the LMS approach in two cases (exp and sin).

Second, in the two cases where the regression functions sin2pi and sin were used, the lowest values for the SD were achieved with the nonparametric approach. Higher values were obtained with the SEMM approach for these functions. The parametric LMS approach showed the second best SD values. In the case where the exponential (exp) regression function was used, the best SD values were obtained with the SEMM approach. The nonparametric approach showed SD values close to the SD values of the SEMM approach. Highest SD values were obtained with the LMS approach.

Third, the nonparametric approach showed the smallest RMSE for all functions (because the difference between the SD for the exponential function was small compared with the bias). In the sin2pi case, the LMS approach showed a slightly better RMSE value than the SEMM approach. The SEMM approach was better than the LMS approach in the exp and sin cases.

4.3.2. Simulation study 2

In this paragraph, we will present the results of the simulation study, in which the amount of non-normality of the latent predictor variable Z_1 (and the sample size) was varied. Furthermore, a second order polynomial model (with quadratic effects) was the true model (see Eq. (16)).

First, with a normally distributed latent predictor Z_1 (skewness 0 and kurtosis 0), Table 2 shows the bias, standard deviation (SD), and root mean square error (RMSE) of the regression function estimates for the three sample sizes ($N = 250, 500$ and 1000). For an easier interpretation the lowest values for each case (in columns) are bolded. The results can be summarized as follows: As can be expected the parametric LMS approach produced unbiased estimates of the quadratic regression function, because all assumptions (correct specifications) of the LMS approach met the true model. Furthermore, LMS showed the lowest RMSE values across the three sample sizes. The nonparametric approach showed slightly better bias and RMSE values than the SEMM approach for all sample sizes. The best SD values were obtained with the SEMM approach. The convergence claimed in Theorems 1 and 2 can be seen in the decreasing RMSE for a rising number of observations for the nonparametric approach.

Second, for a moderate non-normally distributed latent predictor Z_1 (skewness 2 and kurtosis 7), Table 3 shows the bias, SD, and RMSE of the regression function estimates for the three sample sizes. The results can be summarized as follows: The best bias and RMSE values were obtained with the LMS approach followed by the nonparametric approach (for all sample sizes). The SEMM approach produced larger bias and RMSE values. Its SD values were best. The LMS approach and

Table 2
Results of the simulation study with varying degrees of non-normality of Z_1 — skewness 0 and kurtosis 0.

N		250	500	1000
Bias	N_PAR	0.308	0.323	0.320
	AIC	0.716	0.693	0.712
	BIC	0.701	0.718	0.710
	LMS	**0.024**	**0.016**	**0.008**
SD	N_PAR	0.326	0.214	0.133
	AIC	**0.123**	**0.091**	**0.066**
	BIC	0.220	0.091	0.066
	LMS	0.196	0.147	0.097
RMSE	N_PAR	0.448	0.388	0.346
	AIC	0.727	0.699	0.715
	BIC	0.735	0.723	0.713
	LMS	**0.197**	**0.148**	**0.097**

Note. N_PAR = nonparametric approach; AIC = SEMM approach with AIC; BIC = SEMM approach with BIC; LMS = Latent Moderated Structural Equations approach. The lowest values of bias, SD, and RMSE are bolded within each case (column).

Table 3
Results of the simulation study with varying degrees of non-normality of Z_1 — skewness 2 and kurtosis 7.

N		250	500	1000
Bias	N_PAR	0.494	0.453	0.471
	AIC	1.708	1.847	1.838
	BIC	1.635	1.697	1.745
	LMS	**0.410**	**0.392**	**0.449**
SD	N_PAR	0.346	0.226	0.196
	AIC	**0.308**	**0.211**	**0.170**
	BIC	**0.308**	**0.211**	**0.170**
	LMS	0.430	0.232	0.189
RMSE	N_PAR	0.603	0.506	0.510
	AIC	1.732	1.859	1.846
	BIC	1.663	1.710	1.754
	LMS	**0.594**	**0.455**	**0.488**

Note. N_PAR = nonparametric approach; AIC = SEMM approach with AIC; BIC = SEMM approach with BIC; LMS = Latent Moderated Structural Equations approach. The lowest values of bias, SD, and RMSE are bolded within each case (column).

nonparametric approach showed slightly higher SD values. The convergence claimed in Theorems 1 and 2 can be seen since the RMSE decreases if more observations than $N = 250$ are used for the nonparametric approach.

Third, for a strongly non-normally distributed latent predictor Z_1 (skewness 3 and kurtosis 21), Table 4 shows the bias, SD, and RMSE of the regression function estimates for the three sample sizes. The results can be summarized as follows: The nonparametric approach consistently showed lowest bias, SD, and RMSE values for all sample sizes. The LMS approach showed second best bias and RMSE values. The SEMM approach produced better SD values than the LMS approach. Again, the convergence claimed in Theorems 1 and 2 can be seen in the decreasing RMSE for a rising number of observations for the nonparametric approach.

4.3.3. Simulation study 3

In this paragraph, we present the results of the simulation study, in which two-dimensional latent predictor variable Z_1 was chosen. The relation between the latent variables (Z_1 and Z_2) was given by two different regression functions (sin2dim and quad). Table 5 presents the bias, standard deviation (SD), and root mean square error (RMSE) of the regression function estimates for the four different conditions (resulting from Eqs. (19) and (20) as well as from the uniform and normal distribution).

The results can be summarized as follows: First, the nonparametric approach yielded substantially lower bias in all cases than the SEMM approach and the LMS approach. The SEMM approach performed better than the LMS approach when uniform distributions were given. The LMS approach exhibited slightly lower bias than the SEMM approach in the case of the sinusoidal function with normal distribution.

Second, in three cases, the nonparametric approach yielded lower standard deviation than the SEMM approach and the LMS approach. In one case, the LMS approach achieved the lowest standard deviation for the quad (rational) function and

Table 4
Results of the simulation study with varying degrees of non-normality of Z_1 — skewness 3 and kurtosis 21.

N		250	500	1000
Bias	N_PAR	**0.349**	**0.382**	**0.347**
	AIC	2.802	2.699	2.808
	BIC	3.076	2.543	3.451
	LMS	0.948	0.885	1.103
SD	N_PAR	**0.415**	**0.251**	**0.161**
	AIC	0.632	0.563	0.403
	BIC	0.632	0.563	0.403
	LMS	1.372	0.612	1.076
RMSE	N_PAR	**0.542**	**0.457**	**0.383**
	AIC	2.872	2.757	2.837
	BIC	3.131	2.605	3.474
	LMS	1.668	1.076	1.541

Note. N_PAR = nonparametric approach; AIC = SEMM approach with AIC; BIC = SEMM approach with BIC; LMS = Latent Moderated Structural Equations approach. The lowest values of bias, SD, and RMSE are bolded within each case (column).

Table 5
Results of the simulation study with two-dimensional uniformly and normally distributed latent variables and 2 regression functions.

Function distribution		sin2dim		quad	
		Uniform	Normal	Uniform	Normal
Bias	N_PAR	**0.025**	**0.023**	**0.084**	**0.134**
	AIC	0.131	0.962	1.494	0.896
	BIC	0.130	0.961	1.483	0.897
	LMS	0.458	0.936	6.776	1.325
SD	N_PAR	**0.051**	**0.112**	**0.093**	0.132
	AIC	0.052	0.369	0.162	0.342
	BIC	0.053	0.369	0.170	0.341
	LMS	0.096	0.175	0.298	**0.099**
RMSE	N_PAR	**0.057**	**0.115**	**0.126**	**0.188**
	AIC	0.141	1.030	1.503	0.959
	BIC	0.141	1.029	1.493	0.960
	LMS	0.468	0.952	6.782	1.329

Note. N_PAR = nonparametric approach; AIC = SEMM approach with AIC; BIC = SEMM approach with BIC; LMS = Latent Moderated Structural Equations approach. sin2dim represents the sinusoidal function in Eq. (19). quad represents the rational function with quadratic terms in Eq. (20). The lowest values of bias, SD, and RMSE are bolded within each case (column).

normal distribution. In two cases of uniform distributions the SEMM approach had lower standard deviations than the LMS approach. In two cases of normal distributions the LMS approach had lower standard deviations than the SEMM approach.

Third, the nonparametric approach achieved the smallest RMSE in all cases. The SEMM approach was better than the LMS approach in three cases. The LMS approach achieved slightly better RMSE than the SEMM approach for sinusoidal regression function with normal distribution.

To recapitulate the proposed nonparametric approach worked well for multi-dimensional latent predictor variables as well as for one-dimensional latent predictor variables as demonstrated in the two previous simulation studies.

5. Proofs

5.1. Proof of Lemma 1

The proof is an extension of the proof of Lemma 2.1 in Li [27]. Set $a_1 = b_1 = 1 = \tilde{a}_1 = \tilde{b}_1$. For $j, k = 1, \ldots, d, j \neq k$, we have

$$\mathbf{E}(X^{(j)}X^{(k)}) = \mathbf{E}\left\{(a_jZ_1 + \epsilon_j)(a_kZ_1 + \epsilon_k)\right\} = a_j\, a_k\mathbf{E}(Z_1^2),$$

where the last equality follows from the independence assumption and $\mathbf{E}(\epsilon_k) = 0$ for all $k \in \{1, \ldots, d\}$, and similarly

$$\mathbf{E}(X^{(j)}X^{(k)}) = \tilde{a}_j\tilde{a}_k\mathbf{E}(\tilde{Z}_1^2).$$

Since a_2, a_3 and $\mathbf{E}(Z_1^2)$ are nonzero, $\tilde{a}_2$, $\tilde{a}_3$ and $\mathbf{E}(\tilde{Z}_1^2)$ share this property. Hence for $j = 2$ we have

$$a_2 = \frac{a_2 a_3 \mathbf{E}(Z_1^2)}{1 a_3 \mathbf{E}(Z_1^2)} = \frac{\mathbf{E}(X^{(2)} X^{(3)})}{\mathbf{E}(X^{(1)} X^{(3)})} = \frac{\tilde{a}_2 \tilde{a}_3 \mathbf{E}(\tilde{Z}_1^2)}{1 \tilde{a}_3 \mathbf{E}(\tilde{Z}_1^2)} = \tilde{a}_2$$

and for $j = 3, \ldots, d$ we get

$$a_j = \frac{a_2 a_j \mathbf{E}(Z_1^2)}{1 a_2 \mathbf{E}(Z_1^2)} = \frac{\mathbf{E}(X^{(2)} X^{(j)})}{\mathbf{E}(X^{(1)} X^{(2)})} = \frac{\tilde{a}_2 \tilde{a}_j \mathbf{E}(\tilde{Z}_1^2)}{1 \tilde{a}_2 \mathbf{E}(\tilde{Z}_1^2)} = \tilde{a}_j.$$

Similarly we get

$$b_2 = \frac{\mathbf{E}(Y^{(2)} Y^{(3)})}{\mathbf{E}(Y^{(1)} Y^{(3)})} = \tilde{b}_2, \quad \text{and} \quad b_k = \frac{\mathbf{E}(Y^{(2)} Y^{(k)})}{\mathbf{E}(Y^{(1)} Y^{(2)})} = \tilde{b}_k$$

for $k = 3, \ldots, \ell$.

Using (4) and the independence assumption we see that the characteristic function $\varphi_{(X,Y)}$ of (X, Y) is given by

$$\begin{aligned}
\varphi_{(X,Y)}(u_1, \ldots, u_d, v_1, \ldots, v_\ell) &= \mathbf{E}\left\{ \exp\left(i \sum_{j=1}^{d} u_j X^{(j)} + i \sum_{k=1}^{\ell} v_k Y^{(k)} \right) \right\} \\
&= \mathbf{E}\left[\exp\left\{ i \sum_{j=1}^{d} u_j \left(a_j Z_1 + \epsilon_j \right) + i \sum_{k=1}^{\ell} v_k \left(b_k Z_2 + \delta_k \right) \right\} \right] \\
&= \mathbf{E}\left[\exp\left\{ i \left(\sum_{j=1}^{d} u_j a_j Z_1 + \sum_{k=1}^{\ell} v_k b_k Z_2 \right) \right\} \prod_{j=1}^{d} \exp\left(i u_j \epsilon_j \right) \prod_{k=1}^{\ell} \exp\left(i v_k \delta_k \right) \right] \\
&= \varphi_{(Z_1,Z_2)}\left(\sum_{j=1}^{d} u_j a_j, \sum_{k=1}^{\ell} v_k b_k \right) \prod_{j=1}^{d} \varphi_{\epsilon_j}(u_j) \prod_{k=1}^{\ell} \varphi_{\delta_k}(v_k).
\end{aligned}$$

Since we know that the characteristic function of (X, Y) does not vanish at any point, we can conclude that also $\varphi_{(Z_1,Z_2)}$, φ_{ϵ_j} and φ_{δ_k} share this property. Furthermore, using

$$\varphi_{\epsilon_j}(0) = \varphi_{\delta_k}(0) = 1 \quad (j = 2, \ldots, d, k = 2, \ldots, \ell)$$

and $\varphi'_{\epsilon_2}(0) = i\mathbf{E}\epsilon_2 = 0 = \varphi'_{\delta_2}(0)$ we get

$$\begin{aligned}
&\varphi_{(X,Y)}(u_1, 0, \ldots, 0, v_1, 0, \ldots, 0) = \varphi_{(Z_1,Z_2)}(u_1, v_1) \varphi_{\epsilon_1}(u_1) \varphi_{\delta_1}(v_1), \\
&\frac{\partial}{\partial u_2} \varphi_{(X,Y)}(u_1, 0, \ldots, 0, v_1, 0, \ldots, 0) \\
&\quad = a_2 \frac{\partial}{\partial z_1} \varphi_{(Z_1,Z_2)}(u_1, v_1) \varphi_{\epsilon_1}(u_1) \varphi_{\delta_1}(v_1) + \varphi_{(Z_1,Z_2)}(u_1, v_1) \varphi_{\epsilon_1}(u_1) \varphi_{\delta_1}(v_1) \varphi'_{\epsilon_2}(0) \\
&\quad = a_2 \frac{\partial}{\partial z_1} \varphi_{(Z_1,Z_2)}(u_1, v_1) \varphi_{\epsilon_1}(u_1) \varphi_{\delta_1}(v_1)
\end{aligned}$$

and

$$\frac{\partial}{\partial v_2} \varphi_{(X,Y)}(u_1, 0, \ldots, 0, v_1, 0, \ldots, 0) = b_2 \frac{\partial}{\partial z_2} \varphi_{(Z_1,Z_2)}(u_1, v_1) \varphi_{\epsilon_1}(u_1) \varphi_{\delta_1}(v_1).$$

We conclude

$$\begin{aligned}
\varphi_{(Z_1,Z_2)}(u, v) &= \exp\left[\{ \ln \varphi_{(Z_1,Z_2)}(u, v) - \ln \varphi_{(Z_1,Z_2)}(u, 0) \} \right] \exp\left[\{ \ln \varphi_{(Z_1,Z_2)}(u, 0) - \ln \varphi_{(Z_1,Z_2)}(0, 0) \} \right] \\
&= \exp\left\{ \int_0^v \frac{1}{b_2} \frac{\frac{\partial}{\partial v_2} \varphi_{(X,Y)}(u, 0, \ldots, 0, s, 0, \ldots, 0)}{\varphi_{(X,Y)}(u, 0, \ldots, 0, s, 0, \ldots, 0)} \, ds \right\} \\
&\quad \times \exp\left\{ \int_0^u \frac{1}{a_2} \frac{\frac{\partial}{\partial u_2} \varphi_{(X,Y)}(t, 0, \ldots, 0, 0, 0, \ldots, 0)}{\varphi_{(X,Y)}(t, 0, \ldots, 0, 0, 0, \ldots, 0)} \, dt \right\}.
\end{aligned}$$

We have considered the integrals above as parametrization of complex curve integrals of the function $z \mapsto 1/z$ and split them into finitely many integrals such that $\ln z$ is well defined for each integral. (Here the number of intervals is finite since

the curves in the integrals above have finite length and a positive distance to the origin.) This results in additional factor $\exp(is2\pi) = 1$ for some $s \in \mathbb{N}$. Similarly we get

$$\varphi_{(\tilde{Z}_1,\tilde{Z}_2)}(u,v) = \exp\left(\int_0^v \frac{1}{\tilde{b}_2} \frac{\frac{\partial}{\partial v_2}\varphi_{(X,Y)}(u,0,\ldots,0,s,0,\ldots,0)}{\varphi_{(X,Y)}(u,0,\ldots,0,s,0,\ldots,0)}\,ds\right)$$
$$\times \exp\left(\int_0^u \frac{1}{\tilde{a}_2} \frac{\frac{\partial}{\partial u_2}\varphi_{(X,Y)}(t,0,\ldots,0,0,0,\ldots,0)}{\varphi_{(X,Y)}(t,0,\ldots,0,0,0,\ldots,0)}\,dt\right)$$

and from $a_2 = \tilde{a}_2$ and $b_2 = \tilde{b}_2$ we conclude $\varphi_{(Z_1,Z_2)} = \varphi_{(\tilde{Z}_1,\tilde{Z}_2)}$. But from $\varphi_{(Z_1,Z_2)}$ and $a_1,\ldots,a_d, b_1,\ldots,b_\ell$ we can determine φ_{ϵ_j} and φ_{δ_k} via

$$\varphi_{(X,Y)}(0,\ldots,0,u_j,0,\ldots,0,0,\ldots,0) = \varphi_{(Z_1,Z_2)}\left(u_j a_j, 0\right)\varphi_{\epsilon_j}(u_j)$$

and

$$\varphi_{(X,Y)}(0,\ldots,0,0,\ldots,0,v_k,0,\ldots,0) = \varphi_{(Z_1,Z_2)}\left(0, v_k b_k\right)\varphi_{\delta_k}(v_k).$$

Using the same relation for $\varphi_{(\tilde{Z}_1,\tilde{Z}_2)}$, $\varphi_{\tilde{\epsilon}_j}$ and $\varphi_{\tilde{\delta}_k}$ we see that $\varphi_{\epsilon_j} = \varphi_{\tilde{\epsilon}_j}$ and $\varphi_{\delta_k} = \varphi_{\tilde{\delta}_k}$, which implies the assertion. □

5.2. Proof of Theorem 1

Throughout the proof we will use the abbreviation

$$\int f\{(u_1,u_2),v_1,\ldots,v_d,w_1,\ldots,w_\ell\}\,d\hat{\mu}_n^{(\hat{z}_1,\hat{z}_2)_1^n}$$
$$= \int f\{(u_1,u_2),v_1,\ldots,v_d,w_1,\ldots,w_\ell\}\,\hat{\mu}_n^{(\hat{z}_1,\hat{z}_2)_1^n} d((u_1,u_2),v_1,\ldots,v_d,w_1,\ldots,w_\ell),$$

so, e.g.,

$$\int \sigma\{-n(u_1-\alpha_{r,1})\}\sigma\{-n(u_2-\alpha_{r,2})\}d\hat{\mu}_n^{(\hat{z}_1,\hat{z}_2)_1^n} = \frac{1}{n}\sum_{i=1}^n \sigma\{-n(\hat{z}_{1,i}-\alpha_{r,1})\}\sigma\{-n(\hat{z}_{i,2}-\alpha_{r,2})\}$$

and

$$\int v_j\,d\hat{\mu}_n^{(\hat{z}_1,\hat{z}_2)_1^n} = \frac{1}{n}\sum_{i=1}^n \hat{\epsilon}_{j,i}.$$

The proof is divided into nine steps. The outline of the proof is as follows: We will show that for every subsequence $(n_r)_r$ of $(n)_n$ there exists a subsequence $(n_{r_k})_k$ such that

$$\hat{\mu}_{n_{r_k}}^{(\hat{z}_1,\hat{z}_2)_1^{n_{r_k}}} \to \mu \quad \text{weakly.}$$

To do this, we show in the first step of the proof that $(\hat{\mu}_n^{(\hat{z}_1,\hat{z}_2)_1^n})_{n\in\mathbb{N}}$ is tight with probability 1, which implies the existence of a measure $\tilde{\mu}$ satisfying

$$\hat{\mu}_{n_{r_k}}^{(\hat{z}_1,\hat{z}_2)_1^{n_{r_k}}} \to \tilde{\mu} \quad \text{weakly.}$$

We show then in steps 3 till 7 that $\tilde{\mu}$ has properties, which enable us to conclude via Lemma 1 that $\tilde{\mu} = \mu$.

In the first step of the proof we show that $(\hat{\mu}_n^{(\hat{z}_1,\hat{z}_2)_1^n})_{n\in\mathbb{N}}$ is tight with probability 1, i.e., with probability 1 we find for each $\epsilon > 0$ a compact set $K \subseteq \mathbb{R}^2 \times \mathbb{R}^{d+\ell}$ such that

$$\hat{\mu}_n^{(\hat{z}_1,\hat{z}_2)_1^n}(K^c) \le \epsilon \quad \text{for all } n \in \mathbb{N}.$$

By the strong law of large numbers we know that, with probability 1,

$$\frac{1}{n}\sum_{i=1}^n (X_i^{(j)})^2 \to \mathbf{E}\{(X^{(j)})^2\} < \infty \quad \text{and} \quad \frac{1}{n}\sum_{i=1}^n (Y_i^{(k)})^2 \to \mathbf{E}\{(Y^{(k)})^2\} < \infty, \tag{24}$$

so by definition of the estimate we may assume without loss of generality that

$$\frac{1}{n}\sum_{i=1}^n (X_i^{(j)})^2 \le c, \qquad \frac{1}{n}\sum_{i=1}^n (Y_i^{(k)})^2 \le c, \qquad \frac{1}{n}\sum_{i=1}^n \hat{z}_{i,1}^2 \le c \quad \text{and} \quad \frac{1}{n}\sum_{i=1}^n \hat{z}_{i,2}^2 \le c \tag{25}$$

for all $n \in \mathbb{N}$ for some $c > 0$ with probability 1. Furthermore because of

$$\hat{a}_j \to a_j \quad (n \to \infty) \quad \text{and} \quad \hat{b}_k \to b_k \quad (n \to \infty) \tag{26}$$

with probability 1, we may assume in addition that $|\hat{a}_j| \le c$ and $|\hat{b}_k| \le c$ with probability 1. By Markov's inequality we get

$$\begin{aligned}
&\hat{\mu}_n^{(\hat{z}_1,\hat{z}_2)_1^n}\left\{([-M,M]^{2+d+\ell})^c\right\} \\
&\le \hat{\mu}_n^{(\hat{z}_1,\hat{z}_2)_1^n}\{|u_1| > M\} + \hat{\mu}_n^{(\hat{z}_1,\hat{z}_2)_1^n}\{|u_2| > M\} + \sum_{j=1}^{d} \hat{\mu}_n^{(\hat{z}_1,\hat{z}_2)_1^n}\{|v_j| > M\} + \sum_{k=1}^{\ell} \hat{\mu}_n^{(\hat{z}_1,\hat{z}_2)_1^n}\{|w_k| > M\} \\
&\le \frac{\int |u_1|^2 d\hat{\mu}_n^{(\hat{z}_1,\hat{z}_2)_1^n}}{M^2} + \frac{\int |u_2|^2 d\hat{\mu}_n^{(\hat{z}_1,\hat{z}_2)_1^n}}{M^2} + \sum_{j=1}^{d} \frac{\int |v_j|^2 d\hat{\mu}_n^{(\hat{z}_1,\hat{z}_2)_1^n}}{M^2} + \sum_{k=1}^{d} \frac{\int |w_k|^2 d\hat{\mu}_n^{(\hat{z}_1,\hat{z}_2)_1^n}}{M^2} \\
&= \frac{\frac{1}{n}\sum_{i=1}^{n} \hat{z}_{1,i}^2}{M^2} + \frac{\frac{1}{n}\sum_{i=1}^{n} \hat{z}_{2,i}^2}{M^2} + \sum_{j=1}^{d} \frac{\frac{1}{n}\sum_{i=1}^{n} (X_i^{(j)} - \hat{a}_j \hat{z}_{1,i})^2}{M^2} + \sum_{k=1}^{\ell} \frac{\frac{1}{n}\sum_{i=1}^{n} (Y_i^{(k)} - \hat{b}_k \hat{z}_{2,i})^2}{M^2} \\
&\le \frac{c}{M^2} + \frac{c}{M^2} + d\frac{2c + 2c^3}{M^2} + \ell\frac{2c + 2c^3}{M^2} \le \epsilon
\end{aligned}$$

for M sufficiently large.

In the second step of the proof we show

$$T_n \to 0 \quad a.s. \tag{27}$$

Let $\tilde{T}_n$ and $\hat{\mu}_n^{(Z_1,Z_2)_1^n}$ be defined as T_n and $\hat{\mu}_n^{(\hat{z}_1,\hat{z}_2)_1^n}$, respectively, with $(\hat{z}_{i,1}, \hat{z}_{i,2})$ be replaced by $(Z_{1,i}, Z_{2,i})(i = 1, \ldots, n)$. Because of

$$\mathbf{E}\{(X^{(1)})^2\} = \mathbf{E}(Z_1^2) + \mathbf{E}(\epsilon_1^2)$$

we have $\mathbf{E}Z_1^2 \le \mathbf{E}\{(X^{(1)})^2\} < \infty$, so by the strong law of large numbers we get

$$\frac{1}{n}\sum_{i=1}^{n} Z_{1,i}^2 \to \mathbf{E}Z_1^2 \le \mathbf{E}\{(X^{(1)})^2\} = \lim_{n\to\infty} \frac{1}{n}\sum_{i=1}^{n} (X_i^{(1)})^2 \quad a.s.,$$

hence with probability 1 for n large enough

$$\frac{1}{n}\sum_{i=1}^{n} Z_{1,i}^2 \le 1 + \frac{1}{n}\sum_{i=1}^{n} (X_i^{(1)})^2.$$

Similarly we see that with probability 1 we have for n large enough

$$\frac{1}{n}\sum_{i=1}^{n} Z_{2,i}^2 \le 1 + \frac{1}{n}\sum_{i=1}^{n} (Y_i^{(1)})^2.$$

Then by definition of T_n we have with probability 1 for n large enough $T_n \le \tilde{T}_n$, so it suffices to show $\tilde{T}_n \to 0 \quad a.s.$ Since $(p_r)_{r\in\mathbb{N}}$ are probability weights and since σ is bounded this in turn follows from

$$\left(\int v_j d\hat{\mu}_n^{(Z_1,Z_2)_1^n}\right)^2 \to 0 \quad a.s.\ (j = 1, \ldots, d), \tag{28}$$

$$\left(\int w_k d\hat{\mu}_n^{(Z_1,Z_2)_1^n}\right)^2 \to 0 \quad a.s.\ (k = 1, \ldots, \ell) \tag{29}$$

and for any $r \in \mathbb{N}$,

$$\begin{aligned}
&\Bigg| \int \sigma\{-n(u_1 - \alpha_{r,1})\}\sigma\{-n(u_2 - \alpha_{r,2})\} \prod_{j=1}^{d} \sigma\{-n(v_j - \beta_{r,j})\} \prod_{k=1}^{\ell} \sigma\{-n(w_k - \gamma_{r,k})\} d\hat{\mu}_n^{(Z_1,Z_2)_1^n} \\
&\quad - \int \sigma\{-n(u_1 - \alpha_{r,1})\}\sigma\{-n(u_2 - \alpha_{r,2})\} d\hat{\mu}_n^{(Z_1,Z_2)_1^n} \prod_{j=1}^{d} \int \sigma\{-n(v_j - \beta_{r,j})\} d\hat{\mu}_n^{(Z_1,Z_2)_1^n} \\
&\quad \prod_{k=1}^{\ell} \int \sigma\{-n(w_k - \gamma_{r,k})\} d\hat{\mu}_n^{(Z_1,Z_2)_1^n} \Bigg|^2 \to 0 \quad a.s.
\end{aligned} \tag{30}$$

Let $\bar{\mu}_n^{(\hat{z}_1,\hat{z}_2)_1^n}$ and $\bar{\mu}_n^{(Z_1,Z_2)_1^n}$ be the empirical measures which we get if we replace in the definition of $\hat{\mu}_n^{(\hat{z}_1,\hat{z}_2)_1^n}$ and $\hat{\mu}_n^{(Z_1,Z_2)_1^n}$ the estimated coefficients by the true coefficients, respectively. The proof of step 1 implies that $(\bar{\mu}_n^{(\hat{z}_1,\hat{z}_2)_1^n})_{n\in\mathbb{N}}$ and $(\bar{\mu}_n^{(Z_1,Z_2)_1^n})_{n\in\mathbb{N}}$ are tight with probability 1, too. Since the estimated coefficients converge by the strong law of large numbers almost surely to the true coefficients, we conclude that we have for any bounded, uniformly continuous function f

$$\int f\, d\hat{\mu}_n^{(\hat{z}_1,\hat{z}_2)_1^n} - \int f\, d\bar{\mu}_n^{(\hat{z}_1,\hat{z}_2)_1^n} \to 0 \quad a.s. \quad \text{and} \quad \int f\, d\hat{\mu}_n^{(Z_1,Z_2)_1^n} - \int f\, d\bar{\mu}_n^{(Z_1,Z_2)_1^n} \to 0 \quad a.s. \tag{31}$$

Here we have used that because of the tightness of the measures w.l.o.g. we can integrate (31) over some compact set, so that all occurring variables are bounded.

Furthermore, since $\bar{\mu}_n^{(Z_1,Z_2)_1^n}$ is in fact an empirical distribution to independent and identically distributed data, we know again by the strong law of large numbers that we have in addition

$$\int f\, d\bar{\mu}_n^{(Z_1,Z_2)_1^n} \to \int f\, d\mu \quad a.s.,$$

so altogether we know that we have for all bounded, uniformly continuous functions f

$$\int f\, d\hat{\mu}_n^{(Z_1,Z_2)_1^n} \to \int f\, d\mu \quad a.s.$$

Because of our independence assumption, which implies

$$\begin{aligned}
&\mathbf{E}\left[\sigma\{-n(Z_1-\alpha_{r,1})\}\sigma\{-n(Z_2-\alpha_{r,2})\}\prod_{j=1}^{d}\sigma\{-n(\epsilon_j-\beta_{r,j})\}\prod_{k=1}^{\ell}\sigma\{-n(\delta_k-\gamma_{r,k})\}\right]\\
&\quad=\mathbf{E}\left[\sigma\{-n(Z_1-\alpha_{r,1})\}\sigma\{-n(Z_2-\alpha_{r,2})\}\right]\prod_{j=1}^{d}\mathbf{E}\left[\sigma\{-n(\epsilon_j-\beta_{r,j})\}\right]\prod_{k=1}^{\ell}\mathbf{E}\left[\sigma\{-n(\delta_k-\gamma_{r,k})\}\right],
\end{aligned}$$

from this we conclude (30). Relation (28) follows from $\mathbf{E}\epsilon_j = 0$ and the strong law of large numbers, which implies

$$\int v_j\, d\hat{\mu}_n^{(Z_1,Z_2)_1^n} = \frac{1}{n}\sum_{i=1}^{n}(X_i^{(j)} - \hat{a}_j Z_{1,i}) \to \mathbf{E}\{X^{(j)} - a_j Z_1\} = \mathbf{E}\epsilon_j \quad a.s.$$

Similarly we conclude (29) from $\mathbf{E}\delta_k = 0$.

In the third step of the proof we set $S_j(x_1,\dots,x_{2+d+\ell}) = a_j x_1 + x_{j+2}$ for $j \in \{1,\dots,d\}$ and $S_j(x_1,\dots,x_{2+d+\ell}) = b_{j-d}x_2 + x_{j+2}$ for $j \in \{d+1,\dots,d+\ell\}$ and show that we have, with probability 1,

$$\left(\hat{\mu}_n^{(\hat{z}_1,\hat{z}_2)_1^n}\right)_{(S_1,\dots,S_{d+\ell})} \to \mathbf{P}_{(X^{(1)},\dots,X^{(d)},Y^{(1)},\dots,Y^{(\ell)})} \quad \text{weakly.} \tag{32}$$

To see this, we set

$$\bar{\epsilon}_{j,i} = X_i^{(j)} - a_j\hat{z}_{1,i} \quad \text{and} \quad \bar{\delta}_{k,i} = Y_i^{(k)} - b_k\hat{z}_{2,i}$$

and observe that our estimates of the random variables satisfy trivially the equations

$$X_i^{(j)} = a_j\hat{z}_{1,i} + X_i^{(j)} - a_j\hat{z}_{1,i} = S_j(\hat{z}_{1,i},\hat{z}_{2,i},\bar{\epsilon}_{1,i},\dots,\bar{\epsilon}_{d,i},\bar{\delta}_{1,i},\dots,\bar{\delta}_{\ell,i})$$

and

$$Y_i^{(k)} = b_k\hat{z}_{2,i} + Y_i^{(k)} - b_k\hat{z}_{2,i} = S_{d+k}(\hat{z}_{1,i},\hat{z}_{2,i},\hat{\epsilon}_{1,i},\dots,\bar{\epsilon}_{d,i},\bar{\delta}_{1,i},\dots,\bar{\delta}_{\ell,i}),$$

from which we conclude

$$\left(\bar{\mu}_n^{(\hat{z}_1,\hat{z}_2)_1^n}\right)_{(S_1,\dots,S_{d+\ell})} = \mu_{n,(X,Y)_1^n},$$

where the distribution on the right-hand side is the empirical distribution to $(X_1,Y_1),\dots,(X_n,Y_n)$. But this distribution converges weakly to $\mathbf{P}_{(X,Y)}$, and together with (31) and the continuity of $S_1,\dots,S_{d+\ell}$ this implies (32).

In the fourth step of the proof we show that, with probability 1, there exists a subsequence $(n_r)_r$ of $(n)_n$ and a measure μ satisfying

$$\hat{\mu}_{n_r}^{(\hat{z}_1,\hat{z}_2)_1^{n_r}} \to \mu \quad \text{weakly} \tag{33}$$

and

$$\mu_{(S_1,\dots,S_{d+\ell})} = \mathbf{P}_{(X,Y)}. \tag{34}$$

To see this, observe that by the first step of the proof the measures $\hat{\mu}_n^{(\hat{z}_1,\hat{z}_2)_1^n}$ are tight, and hence according to the theorem of Prohorov (see, e.g., Theorem 6.1 in [63]) relatively compact, so (33) holds. Since $S_1, \ldots, S_{d+\ell}$ are continuous, this implies

$$\left(\hat{\mu}_{n_r}^{(\hat{z}_1,\hat{z}_2)_1^{n_r}}\right)_{(S_1,\ldots,S_{d+\ell})} \to \mu_{(S_1,\ldots,S_{d+\ell})} \quad \text{weakly},$$

from which we get (34) by (32) and the uniqueness of the limit distribution in the case of weak convergence.

In the fifth step of the proof we show by an approximation of indicator functions of intervals by suitable neural networks that because of (27) the components of μ corresponding to (Z_1, Z_2), $\epsilon_1, \ldots, \epsilon_d$, $\delta_1, \ldots, \delta_\ell$ are independent with probability 1. Let F be the cumulative distribution function of μ, i.e.,

$$F\{(x_1, x_2), e_1, \ldots, e_d, d_1, \ldots, d_\ell\} = \mu\{u_1 \le x_1, u_2 \le x_2, v_1 \le e_1, \ldots, v_d \le e_d, w_1 \le d_1, \ldots, w_\ell \le d_\ell\},$$

and set $F_{(Z_1,Z_2)}(x_1, x_2) = \mu\{u_1 \le x_1, u_2 \le x_2\}$, $F_{\epsilon_j}(e_j) = \mu\{v_j \le e_j\}$ and $F_{\delta_k}(d_k) = \mu\{w_k \le d_k\}$. We have to show that

$$F\{(x_1, x_2), e_1, \ldots, e_d, d_1, \ldots, d_\ell\} = F_{(Z_1,Z_2)}(x_1, x_2) \prod_{j=1}^{d} F_{\epsilon_j}(e_j) \prod_{k=1}^{\ell} F_{\delta_k}(d_k) \tag{35}$$

for all $x_1, x_2, e_1, \ldots, e_d, d_1, \ldots, d_\ell \in \mathbb{R}$.

Since distribution functions are right continuous, it suffices to show (35) for $x_1, x_2, e_1, \ldots, e_d, d_1, \ldots, d_\ell$ in some dense subset of $\mathbb{R}$, which we choose as

$$D = \mathbb{R} \setminus \left\{ x \in \mathbb{R} : \mu\{u_1 = x\} + \mu\{u_2 = x\} + \sum_{j=1}^{d} \mu\{v_j = x\} + \sum_{k=1}^{\ell} \mu\{w_k = x\} > 0 \right\}$$

(which is dense in $\mathbb{R}$ since $\{\ldots\}$ is countable).

Let $x_1, x_2, e_1, \ldots, e_d, d_1, \ldots, d_\ell \in D$. For any $x \in \mathbb{R}$ and any $\epsilon > 0$ we can find $\alpha \in \mathbb{Q}$ satisfying for sufficiently large n

$$-n(z - \alpha) \quad \text{is sufficiently large for } z < x - \epsilon$$

and

$$-n(z - \alpha) \quad \text{is sufficiently small for } z > x - \epsilon$$

such that

$$\left|1_{(-\infty,x]}(z) - \sigma\{-n(z - \alpha)\}\right| \le \epsilon$$

for $z < x - \epsilon$ or $z > x + \epsilon$ in case n sufficiently large. Furthermore, for any $x_1, x_2 \in \mathbb{R}$ and any $\epsilon > 0$ we can find $\alpha_1, \alpha_2 \in \mathbb{Q}$ satisfying

$$\left|1_{(-\infty,x_1]\times(-\infty,x_2]}(z_1, z_2) - \sigma\{-n(z_1 - \alpha_1)\}\sigma\{-n(z_2 - \alpha_2)\}\right| \le \epsilon \tag{36}$$

in case that $z_1 < x_1 - \epsilon$ or $z_1 > x_1 + \epsilon$, and that $z_2 < x_2 - \epsilon$ or $z_2 > x_2 + \epsilon$, for n sufficiently large. To see this, fix $x_1, x_2 \in \mathbb{R}$ and $\epsilon > 0$. Choose $\alpha_1, \alpha_2 \in \mathbb{Q}$ such that

$$\left|1_{(-\infty,x_1]}(z) - \sigma\{-n(z - \alpha_1)\}\right| \le \frac{\epsilon}{2}$$

for $z < x_1 - \epsilon$ or $z > x_1 + \epsilon$, and such that

$$\left|1_{(-\infty,x_2]}(z) - \sigma\{-n(z - \alpha_2)\}\right| \le \frac{\epsilon}{2}$$

for $z < x_2 - \epsilon$ or $z > x_2 + \epsilon$. Then it is easy to see that (36) holds if one considers separately the four cases $z_1 < x_1 - \epsilon$ and $z_2 < x_2 - \epsilon$, $z_1 > x_1 + \epsilon$ and $z_2 < x_2 - \epsilon$, $z_1 < x_1 - \epsilon$ and $z_2 > x_2 + \epsilon$, and $z_1 > x_1 - \epsilon$ and $z_2 > x_2 - \epsilon$.

Consequently for suitably chosen r we see by expanding the terms below in a telescoping sum that we have

$$\begin{aligned}
&\Bigg| F\{(x_1, x_2), e_1, \ldots, e_d, d_1, \ldots, d_\ell\} \\
&\quad - \int \sigma\{-n(u_1 - \alpha_{r,1})\}\sigma\{-n(u_2 - \alpha_{r,2})\} \prod_{j=1}^{d} \sigma\{-n(v_j - \beta_{r,j})\} \prod_{k=1}^{\ell} \sigma\{-n(w_k - \gamma_{r,k})\} d\mu \Bigg| \\
&\le (d + \ell + 1)\epsilon + \mu\{x_1 - \epsilon \le z_1 \le x_1 + \epsilon\} + \mu\{x_2 - \epsilon \le z_2 \le x_2 + \epsilon\} \\
&\quad + \sum_{j=1}^{d} \mu\{e_j - \epsilon \le v_j \le e_j + \epsilon\} + \sum_{k=1}^{\ell} \mu\{d_k - \epsilon \le w_k \le d_k + \epsilon\}
\end{aligned}$$

and

$$\begin{aligned}
&\Bigg| F_{(Z_1,Z_2)}(x_1,x_2)\prod_{j=1}^{d}F_{\epsilon_j}(e_j)\prod_{k=1}^{\ell}F_{\delta_k}(d_k)\\
&\quad-\int\sigma\{-n(u_1-\alpha_{r,1})\}\sigma\{-n(u_2-\alpha_{r,2})\}\,d\mu\prod_{j=1}^{d}\int\sigma\{-n(v_j-\beta_{r,j})\}\,d\mu\prod_{k=1}^{\ell}\int\sigma\{-n(w_k-\gamma_{r,k})\}\,d\mu\Bigg|\\
&\le(d+\ell+1)\epsilon+\mu\{x_1-\epsilon\le z_1\le x_1+\epsilon\}+\mu\{x_2-\epsilon\le z_2\le x_2+\epsilon\}\\
&\quad+\sum_{j=1}^{d}\mu\{e_j-\epsilon\le v_j\le e_j+\epsilon\}+\sum_{k=1}^{\ell}\mu\{d_k-\epsilon\le w_k\le d_k+\epsilon\}.
\end{aligned}$$

For $x_1, x_2, e_1, \ldots, e_d, d_1, \ldots, d_\ell \in D$ the right-hand side above converges to zero for $\epsilon \to 0$, so it suffices to show that we have for any r

$$\begin{aligned}
&\int\sigma\{-n(u_1-\alpha_{r,1})\}\sigma\{-n(u_2-\alpha_{r,2})\}\prod_{j=1}^{d}\sigma\{-n(v_j-\beta_{r,j})\}\prod_{k=1}^{\ell}\sigma\{-n(w_k-\gamma_{r,k})\}d\mu\\
&=\int\sigma\{-n(u_1-\alpha_{r,1})\}\sigma\{-n(u_2-\alpha_{r,2})\}\,d\mu\prod_{j=1}^{d}\int\sigma\{-n(v_j-\beta_{r,j})\}\,d\mu\prod_{k=1}^{\ell}\int\sigma\{-n(w_k-\gamma_{r,k})\}\,d\mu.
\end{aligned}$$

But this in turn follows from (33), since

$$\begin{aligned}
&\int\sigma\{-n(u_1-\alpha_{r,1})\}\sigma\{-n(u_2-\alpha_{r,2})\}\prod_{j=1}^{d}\sigma\{-n(v_j-\beta_{r,j})\}\prod_{k=1}^{\ell}\sigma\{-n(w_k-\gamma_{r,k})\}d\mu\\
&\quad-\int\sigma\{-n(u_1-\alpha_{r,1})\}\sigma\{-n(u_2-\alpha_{r,2})\}\,d\mu\prod_{j=1}^{d}\int\sigma\{-n(v_j-\beta_{r,j})\}\,d\mu\prod_{k=1}^{\ell}\int\sigma\{-n(w_k-\gamma_{r,k})\}\,d\mu\\
&=\lim_{l\to\infty}\Bigg[\int\sigma\{-n(u_1-\alpha_{r,1})\}\sigma\{-n(u_2-\alpha_{r,2})\}\prod_{j=1}^{d}\sigma\{-n(v_j-\beta_{r,j})\}\prod_{k=1}^{\ell}\sigma\{-n(w_k-\gamma_{r,k})\}d\hat{\mu}_{n_\ell}^{(\hat{z}_1,\hat{z}_2)_1^{n_\ell}}\\
&\quad-\int\sigma\{-n(u_1-\alpha_{r,1})\}\sigma\{-n(u_2-\alpha_{r,2})\}\,d\hat{\mu}_{n_\ell}^{(\hat{z}_1,\hat{z}_2)_1^{n_\ell}}\\
&\quad\prod_{j=1}^{d}\int\sigma\{-n(v_j-\beta_{r,j})\}\,d\hat{\mu}_{n_\ell}^{(\hat{z}_1,\hat{z}_2)_1^{n_\ell}}\prod_{k=1}^{\ell}\int\sigma\{-n(w_k-\gamma_{r,k})\}d\hat{\mu}_{n_\ell}^{(\hat{z}_1,\hat{z}_2)_1^{n_\ell}}\Bigg]\\
&=0\quad a.s.
\end{aligned}$$

by (27) and $N_n \to \infty (n \to \infty)$.

In the sixth step of the proof we show that the components of μ are, with probability 1, in L_1. By the Portmanteau theorem (see [63]) and $\hat{\mu}_{n_r}^{(\hat{z}_1,\hat{z}_2)_1^{n_r}} \to \mu$ weakly, with probability 1, we have, with probability 1,

$$\begin{aligned}
\int|u_1|\,d\mu&=\int_0^\infty\mu\{|u_1|>t\}\,dt\\
&\le\int_0^\infty\liminf_{r\to\infty}\hat{\mu}_{n_r}^{(\hat{z}_1,\hat{z}_2)_1^{n_r}}\{|u_1|>t\}\,dt\\
&\le\int_0^\infty\liminf_{r\to\infty}\frac{\int|u_1|^2d\hat{\mu}_{n_r}^{(\hat{z}_1,\hat{z}_2)_1^{n_r}}}{t^2}\,dt<\infty,
\end{aligned}$$

since by definition of the estimate we have, with probability 1,

$$\begin{aligned}
\liminf_{r\to\infty}\int|u_1|^2d\hat{\mu}_{n_r}^{(\hat{z}_1,\hat{z}_2)_1^{n_r}}&=\liminf_{r\to\infty}\frac{1}{n_r}\sum_{i=1}^{n_r}\hat{z}_{1,i}^2\\
&\le\liminf_{r\to\infty}\left\{1+\frac{1}{n_r}\sum_{i=1}^{n_r}(X_i^{(1)})^2\right\}\\
&=1+\mathbf{E}\{(X_i^{(1)})^2\}<\infty.
\end{aligned}$$

Furthermore

$$\int |v_j|\, d\mu \le \int_0^\infty \liminf_{r\to\infty} \frac{\int |v_j|^2 d\hat{\mu}_{n_r}^{(\hat{z}_1,\hat{z}_2)_1^{n_r}}}{t^2}\, dt < \infty \quad a.s.,$$

since we have, with probability 1,

$$\begin{aligned}\liminf_{r\to\infty} \int |v_j|^2 d\hat{\mu}_{n_r}^{(\hat{z}_1,\hat{z}_2)_1^{n_r}} &= \liminf_{r\to\infty} \frac{1}{n_r}\sum_{i=1}^{n_r}\left(X_i^{(j)} - \hat{a}_j \hat{z}_{1,i}\right)^2 \\ &\le \liminf_{r\to\infty}\left\{2\frac{1}{n_r}\sum_{i=1}^{n_r}\left(X_i^{(j)}\right)^2 + 2\hat{a}_j^2 \frac{1}{n_r}\sum_{i=1}^{n_r}\hat{z}_{1,i}^2\right\} \\ &= 2\mathbf{E}\{(X_i^{(1)})^2\} + 2a_j^2(1+\mathbf{E}\{(X_i^{(1)})^2\}) < \infty.\end{aligned}$$

Similar arguments for the other components yield the desired result.

In the seventh step of the proof we show that we have, with probability 1,

$$\int v_j\, d\mu = \int w_k\, d\mu = 0 \quad \text{for } j \in \{1,\dots,d\} \text{ and } k \in \{1,\dots,\ell\}. \tag{37}$$

To do this, we observe that because of (27) we have, with probability 1,

$$\int v_j\, d\hat{\mu}_n^{(\hat{z}_1,\hat{z}_2)_1^n} \to 0 \quad (n\to\infty) \quad \text{and} \quad \int w_k\, d\hat{\mu}_n^{(\hat{z}_1,\hat{z}_2)_1^n} \to 0 \quad (n\to\infty)$$

for $j \in \{1,\dots,d\}$ and $k \in \{1,\dots,\ell\}$. Using the arguments of the sixth step of the proof we see that we have

$$\int |x_j| 1_{\{|x_j|>L\}} d\mu(x_j) \to 0 \quad (L\to\infty)$$

and

$$\int |x_j| 1_{\{|x_j|>L\}} d\hat{\mu}_{n_r}^{(\hat{z}_1,\hat{z}_2)_1^{n_r}}(x_j) \le \frac{1}{L}\int |x_j|^2 d\hat{\mu}_{n_r}^{(\hat{z}_1,\hat{z}_2)_1^{n_r}}(x_j) \to 0 \quad (L\to\infty).$$

Consequently we may replace $(x_1,\dots,x_{2+d+\ell}) \mapsto x_j$ by a bounded and continuous function in the integrals below, hence $\hat{\mu}_{n_r}^{(\hat{z}_1,\hat{z}_2)_1^{n_r}} \to \mu$ weakly implies

$$\int x_j\, d\mu(x_j) = \lim_{r\to\infty} \int x_j\, d\hat{\mu}_{n_r}^{(\hat{z}_1,\hat{z}_2)_1^{n_r}}(x_j) = 0.$$

In the eighth step of the proof we show that we have, with probability 1,

$$\mu = \mathbf{P}_{((Z_1,Z_2),\epsilon^{(1)},\dots,\epsilon^{(d)},\delta^{(1)},\dots,\delta^{(\ell)})}. \tag{38}$$

This follows directly of the uniqueness of the distribution of $((Z_1,Z_2),\epsilon^{(1)},\dots,\epsilon^{(d)},\delta^{(1)},\dots,\delta^{(\ell)})$ shown in Lemma 1 and the properties of the distribution μ proven in the previous four steps.

In the ninth and final step of the proof we show the assertion of the theorem.

Let f be an arbitrary bounded and continuous function. We have to show that, with probability 1, for all such functions

$$\int f\, d\hat{\mu}_n^{(\hat{z}_1,\hat{z}_2)_1^n} \to \int f\, d\mathbf{P}_{((Z_1,Z_2),\epsilon^{(1)},\dots,\epsilon^{(d)},\delta^{(1)},\dots,\delta^{(\ell)})} \quad (n\to\infty).$$

To show this, it suffices to show that, with probability 1, for any subsequence $(n_r)_r$ of $(n)_n$ and all such functions there exists a subsubsequence $(n_{r_k})_k$ with the property

$$\int f\, d\hat{\mu}_{n_{r_k}}^{(\hat{z}_1,\hat{z}_2)_1^{n_{r_k}}} \to \int f\, d\mathbf{P}_{((Z_1,Z_2),\epsilon^{(1)},\dots,\epsilon^{(d)},\delta^{(1)},\dots,\delta^{(\ell)})} \quad (k\to\infty). \tag{39}$$

Let $(n_r)_r$ be an arbitrary subsequence of $(n)_n$. According to steps 1 till 8 above applied to $(n_r)_r$ instead of $(n)_n$ there exists a subsequence $(n_{r_k})_k$ of $(n_r)_r$ with the property

$$\hat{\mu}_{n_{r_k}}^{(\hat{z}_1,\hat{z}_2)_1^{n_{r_k}}} \to \mathbf{P}_{((Z_1,Z_2),\epsilon^{(1)},\dots,\epsilon^{(d)},\delta^{(1)},\dots,\delta^{(\ell)})} \quad \text{weakly.}$$

Here the weak convergence holds whenever (24)–(26) hold. But this implies (39), and the proof is complete. □

5.3. Proof of Theorem 2

Choose $f_n \in \mathcal{F}_n$ such that

$$\int |f_n(z) - m(z)|^2 \mathbf{P}_{Z_1}(dz) \to 0 \quad (n \to \infty).$$

Then

$$\begin{aligned} 0 &\le \int |m_n(z) - m(z)|^2 \mathbf{P}_{Z_1}(dz) \\ &= \int |m_n(z_1) - z_2|^2 d\mu - \int |m(z_1) - z_2|^2 d\mu \\ &= \int |m_n(z_1) - z_2|^2 d\mu - \int |f_n(z_1) - z_2|^2 d\mu + \int |f_n(z) - m(z)|^2 \mathbf{P}_{Z_1}(dz). \end{aligned}$$

Hence it suffices to show

$$\limsup_{n\to\infty} \int |m_n(z_1) - z_2|^2 d\mu - \int |f_n(z_1) - z_2|^2 d\mu \le 0 \quad a.s.$$

Since by definition of m_n

$$\begin{aligned} \int |m_n(z_1) - z_2|^2 d\mu - \int |f_n(z_1) - z_2|^2 d\mu \le \int |m_n(z_1) - z_2|^2 d\mu - \frac{1}{n}\sum_{i=1}^n |m_n(\hat{z}_{i,1}) - \hat{z}_{i,2}|^2 \\ + \frac{1}{n}\sum_{i=1}^n |f_n(\hat{z}_{i,1}) - \hat{z}_{i,2}|^2 - \int |f_n(z_1) - z_2|^2 d\mu \end{aligned}$$

this in turn follows from

$$\int |m_n(z_1) - z_2|^2 d\mu - \frac{1}{n}\sum_{i=1}^n |m_n(\hat{z}_{i,1}) - \hat{z}_{i,2}|^2 \to 0 \quad a.s. \tag{40}$$

and

$$\frac{1}{n}\sum_{i=1}^n |f_n(\hat{z}_{i,1}) - \hat{z}_{i,2}|^2 - \int |f_n(z_1) - z_2|^2 d\mu \to 0 \quad a.s. \tag{41}$$

For $\beta > 0$ and $z \in \mathbb{R}$ set $T_\beta z = \max\{\min\{z, \beta\}, -\beta\}$. We have

$$\int |z_2|^2 1_{\{|z_2|>\beta\}} d\mu \to 0 \quad (\beta \to \infty)$$

by dominated convergence and

$$\begin{aligned} \limsup_{n\to\infty} \frac{1}{n}\sum_{i=1}^n |\hat{z}_{i,2}|^2 1_{\{|\hat{z}_{i,2}|>\beta\}} &\le \limsup_{n\to\infty} \frac{1}{\beta^2}\frac{1}{n}\sum_{i=1}^n |\hat{z}_{i,2}|^4 \\ &\le \frac{1}{\beta^2} \limsup_{n\to\infty} \left\{ 1 + 256\frac{1}{n}\sum_{i=1}^n (Y_i^{(1)})^4 + 272\left(\frac{1}{n}\sum_{i=1}^n Y_i^{(1)}\right)^4 \right\} \\ &\to 0 \quad (\beta \to \infty) \end{aligned}$$

a.s. by (6) and the Strong Law of Large Numbers. Hence in order to prove (40) it suffices to show

$$\int |m_n(z_1) - T_\beta z_2|^2 d\mu - \frac{1}{n}\sum_{i=1}^n |m_n(\hat{z}_{i,1}) - T_\beta \hat{z}_{i,2}|^2 \to 0 \quad a.s.$$

for all $\beta > 0$.

Let $\beta > 0$ be arbitrary. It suffices to show that with probability 1, any subsequence $(n_k)_k$ from $(n)_n$ contains a subsubsequence n_{k_r} such that

$$\int |m_{n_{k_r}}(z_1) - T_\beta z_2|^2 d\mu - \frac{1}{n_{k_r}}\sum_{i=1}^{n_{k_r}} |m_{n_{k_r}}(\hat{z}_{i,1}) - T_\beta \hat{z}_{i,2}|^2 \to 0 \quad (r \to \infty).$$

In the sequel we condition on the event that

$$\hat{\mu}_n^{(\hat{z}_1,\hat{z}_2)_1^n} \to \mu \quad \text{weakly}, \tag{42}$$

which has probability 1 because of Theorem 1. Let $(n_k)_k$ be an arbitrary subsequence of $(n)_n$. By the Arzela–Ascoli Theorem (see [64]) the sequence m_{n_k} of equicontinuous functions contains a (random) subsequence $m_{n_{k_r}}$ which converges in supremum norm to some (random) function $\bar{m}$. Since the functions $m_{n_{k_r}}$ are continuous and bounded, $\bar{m}$ has this property, too. By (42) we know

$$\int |\bar{m}(z_1) - T_\beta z_2|^2 d\mu - \frac{1}{n_{k_r}} \sum_{i=1}^{n_{k_r}} |\bar{m}(\hat{z}_{i,1}) - T_\beta \hat{z}_{i,2}|^2 \to 0 \quad (r \to \infty).$$

Using

$$\left| \int |m_{n_{k_r}}(z_1) - T_\beta z_2|^2 d\mu - \int |\bar{m}(z_1) - T_\beta z_2|^2 d\mu \right| = \left| \int \{m_{n_{k_r}}(z_1) - \bar{m}(z_1)\}\{m_{n_{k_r}}(z_1) + \bar{m}(z_1) - 2T_\beta z_2\} d\mu \right|$$
$$\leq (2L + 2\beta) \|m_{n_{k_r}} - \bar{m}\|_\infty$$

and

$$\left| \frac{1}{n_{k_r}} \sum_{i=1}^{n_{k_r}} |m_{n_{k_r}}(\hat{z}_{i,1}) - T_\beta \hat{z}_{i,2}|^2 - \frac{1}{n_{k_r}} \sum_{i=1}^{n_{k_r}} |\bar{m}(\hat{z}_{i,1}) - T_\beta \hat{z}_{i,2}|^2 \right| \leq (2L + 2\beta) \|m_{n_{k_r}} - \bar{m}\|_\infty,$$

we see that this implies (40). In the same way we can also prove (41), which completes the proof. □

Acknowledgments

The authors would like to thank the referee, an Associate Editor and the Editor-in-Chief for their constructive and insightful comments. They would also like to thank Kenneth A. Bollen for helpful comments on an earlier version of the manuscript. The research of the first and last authors was supported by the Deutsche Forschungsgemeinschaft (DFG) Grant No. KE 1664/1–2. The research of the third author was funded by the Natural Sciences and Engineering Research Council of Canada (NSERC) under Grant RGPIN–2015–06412.

References

[1] K.A. Bollen, Latent variables in psychology and the social sciences, Ann. Rev. Psychol. 53 (2002) 605–634.
[2] A. Skrondal, S. Rabe-Hesketh, Generalized Latent Variable Modeling: Multilevel, Longitudinal and Structural Equation Models, Chapman & Hall/CRC, Boca Raton, FL, 2004.
[3] T. Hastie, R. Tibshirani, J. Friedman, The Elements of Statistical Learning, second ed., Springer-Verlag, New York, 2009.
[4] A. Montanari, C. Viroli, The independent factor analysis approach to latent variable modelling, Statistics 44 (2010) 397–416.
[5] E.S. Allman, C. Matias, J.A. Rhodes, Identifiability of parameters in latent structure models with many observed variables, Ann. Statist. 37 (2009) 3099–3132.
[6] J.B. Kruskal, More factors than subjects, tests and treatments: An indeterminacy theorem for canonical decomposition and individual differences scaling, Psychometrika 41 (1976) 281–293.
[7] J.B. Kruskal, Three-way arrays: Rank and uniqueness of trilinear decompositions, with application to arithmetic complexity and statistics, Linear Algebra Appl. 18 (1977) 95–138.
[8] U. Amato, A. Antoniadis, A. Samarov, A.B. Tsybakov, Noisy independent factor analysis model for density estimation and classification, Electron. J. Stat. 4 (2010) 707–736.
[9] T.W. Anderson, Linear latent variable models and covariance structures, J. Econometrics 41 (1989) 91–119.
[10] J.A. Breslaw, J. McIntosh, Simulated latent variable estimation of models with ordered categorical data, J. Econometrics 87 (1998) 25–47.
[11] M. Gebregziabher, S.M. DeSantis, Latent class based multiple imputation approach for missing categorical data, J. Statist. Plann. Inference 140 (2010) 3252–3262.
[12] J. Bai, S. Ng, Determining the number of factors in approximate factor models, Econometrica 70 (2002) 191–221.
[13] S. Bianconcini, S. Cagnone, Estimation of generalized linear latent variable models via fully exponential Laplace approximation, J. Multivariate Anal. 112 (2012) 183–193.
[14] I. Irincheeva, E. Cantoni, M.G. Genton, Generalized linear latent variable models with flexible distribution of latent variables, Scand. J. Stat. 39 (2012) 1–18.
[15] F. Bartolucci, F. Pennoni, B. Francis, Likelihood inference for a class of latent Markov models under linear hypothesis on the transition probabilities, J. R. Stat. Soc. Ser. B 68 (2006) 155–178.
[16] F. Bartolucci, A latent Markov model for detecting patterns of criminal activity, J. R. Stat. Soc. Ser. A 170 (2007) 115–132.
[17] R.P. Browne, P.D. McNicholas, Model-based clustering, classification, and discriminant analysis of data with mixed type, J. Statist. Plann. Inference 142 (2012) 2976–2984.
[18] P.D. McNicholas, Model-based classification using latent Gaussian mixture models, J. Statist. Plann. Inference 140 (2010) 1175–1181.
[19] W.F. Christensen, Y. Amemiya, Latent variable analysis of multivariate spatial data, J. Amer. Statist. Assoc. 97 (2002) 302–317.
[20] D. Colombo, M.H. Maathuis, M. Kalisch, T.S. Richardson, Learning high-dimensional directed acyclic graphs with latent and selection variables, Ann. Statist. 40 (2012) 294–321.
[21] P. Hall, H.G. Müller, F. Yao, Modelling sparse generalized longitudinal observations with latent Gaussian processes, J. R. Stat. Soc. Ser. B 70 (2008) 703–723.
[22] H.S. Lynn, C.E. McCulloch, Using principal component analysis and correspondence analysis for estimation in latent variable models, J. Amer. Statist. Assoc. 95 (2000) 561–572.
[23] J. Bai, S. Ng, Evaluating latent and observed factors in microeconomics and finance, J. Econometrics 131 (2006) 507–537.
[24] R. Schumacker, G. Marcoulides, Interaction and Nonlinear Effects in Structural Equation Modeling, Lawrence Erlbaum Associates, Hillsdale, NJ, 1998.

[25] D. Paul, E. Bair, T. Hastie, R. Tibshirani, "Preconditioning" for feature selection and regression in high-dimensional problems, Ann. Statist. 36 (2008) 1595–1618.
[26] D. Connes, E. Ronchetti, M.P. Victoria-Feser, Goodness of fit for generalized linear latent variables models, J. Amer. Statist. Assoc. 105 (2010) 1126–1134.
[27] T. Li, Robust and consistent estimation of nonlinear errors-in-variables models, J. Econometrics 110 (2002) 1–26.
[28] L. Devroye, A. Krzyżak, An equivalence theorem for L_1 convergence of the kernel regression estimate, J. Statist. Plann. Inference 23 (1989) 71–82.
[29] L. Devroye, T.J. Wagner, Distribution-free consistency results in nonparametric discrimination and regression function estimation, Ann. Statist. 8 (1980) 231–239.
[30] E.A. Nadaraya, On estimating regression, Theory Probab. Appl. 9 (1964) 141–142.
[31] E.A. Nadaraya, Remarks on nonparametric estimates for density functions and regression curves, Theory Probab. Appl. 15 (1970) 134–137.
[32] C.J. Stone, Consistent nonparametric regression, Ann. Statist. 5 (1977) 595–645.
[33] G.S. Watson, Smooth regression analysis, Sankhyā A 26 (1964) 359–372.
[34] J. Beirlant, L. Györfi, On the asymptotic L_2-error in partitioning regression estimation, J. Statist. Plann. Inference 71 (1998) 93–107.
[35] L. Györfi, Recent results on nonparametric regression estimate and multiple classification, Probl. Control Inf. Theory 10 (1981) 43–52.
[36] L. Devroye, Necessary and sufficient conditions for the almost everywhere convergence of nearest neighbor regression function estimates, Z. Wahrscheinlichkeitstheor. Verwandte Geb. 61 (1982) 467–481.
[37] L. Devroye, L. Györfi, A. Krzyżak, G. Lugosi, On the strong universal consistency of nearest neighbor regression function estimates, Ann. Statist. 22 (1994) 1371–1385.
[38] Y.P. Mack, Local properties of k-nearest neighbor regression estimates, SIAM J. Algebr. Discrete Methods 2 (1981) 311–323.
[39] L.C. Zhao, Exponential bounds of mean error for the nearest neighbor estimates of regression functions, J. Multivariate Anal. 21 (1987) 168–178.
[40] G. Lugosi, K. Zeger, Nonparametric estimation via empirical risk minimization, IEEE Trans. Inform. Theory 41 (1995) 677–687.
[41] M. Kohler, A. Krzyżak, Nonparametric regression estimation using penalized least squares, IEEE Trans. Inform. Theory 47 (2001) 3054–3058.
[42] L. Györfi, M. Kohler, A. Krzyżak, H. Walk, A Distribution-free Theory of Nonparametric Regression, Springer, New York, 2002.
[43] L.L. Thurstone, Multiple-factor Analysis, The University of Chicago Press, Chicago, IL, 1947.
[44] R.B. Cattell, The Scientific Use of Factor Analysis, Plenum, New York, 1978.
[45] R.I. Jennrich, P.F. Sampson, Rotation for simple loadings, Psychometrika 31 (1966) 313–323.
[46] T.A. Brown, Confirmatory Factor Analysis for Applied Research, second ed., The Guilford Press, New York, 2015.
[47] D.J. Bauer, A semiparametric approach to modeling nonlinear relations among latent variables, Struct. Equ. Model. 12 (2005) 513–535.
[48] A.G. Klein, H. Moosbrugger, Maximum likelihood estimation of latent interaction effects with the LMS method, Psychometrika 65 (2000) 457–474.
[49] C. de Boor, A Practical Guide to Splines, Springer-Verlag, New York, 1978.
[50] L. Schumaker, Spline Functions: Basic Theory, Wiley, New York, 1981.
[51] C.J. Stone, Additive regression and other nonparametric models, Ann. Statist. 13 (1985) 689–705.
[52] C.J. Stone, The use of polynomial splines and their tensor products in multivariate function estimation, Ann. Statist. 22 (1994) 118–184.
[53] L. Dempster, N.M. Laird, D.B. Rubin, Maximum likelihood from incomplete data via the EM algorithm, J. R. Stat. Soc. Ser. B Stat. Methodol. 39 (1977) 1–38.
[54] A. Kelava, C.S. Werner, K. Schermelleh-Engel, D. Zapf, H. Moosbrugger, Y. Ma, H. Cham, L.S. Aiken, S.G. West, Advanced nonlinear latent variable modeling: Distribution analytic LMS and QML estimators of interaction and quadratic effects, Struct. Equ. Model. 18 (2011) 465–491.
[55] H. Brandt, A. Kelava, A.G. Klein, A simulation study comparing recent approaches for the estimation of nonlinear effects in SEM under the condition of non-normality, Struct. Equ. Model. 21 (2014) 181–195.
[56] L.K. Muthén, B.O. Muthén, Mplus User's Guide, sixth ed., Muthén and Muthén, Los Angeles, 1998.
[57] D.J. Bauer, R.E. Baldasaro, N.C. Gottfredson, Diagnostic procedures for detecting nonlinear relationships between latent variables, Struct. Equ. Model. 19 (2012) 157–177.
[58] J. Pek, D. Losardo, D.J. Bauer, Confidence intervals for a semiparametric approach to modeling nonlinear relations among latent variables, Struct. Equ. Model. 18 (2011) 537–553.
[59] R.E. Baldasaro, D.J. Bauer, Abstract: Comparing semiparametric and parametric methods for modeling interactions among latent variables, Multivariate Behav. Res. 30 (2011) 1007–1008.
[60] J. Nocedal, S.J. Wright, Numerical Optimization, Springer, New York, 2006.
[61] C. Vale, V. Maurelli, Simulating multivariate nonnormal distributions, Psychometrika 48 (1983) 465–471.
[62] P.J. Curran, S.G. West, J.F. Finch, The robustness of test statistics to nonnormality and specification error in confirmatory factor analysis, Psychol. Methods 1 (1996) 16–29.
[63] P. Billingsley, Convergence of Probability Measures, Wiley, New York, 1968.
[64] N. Dunford, T.J. Schwartz, Linear Operators, Vol. 1, Wiley, New York, 1958.